

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 17:59:34 ; Search time 7262.23 Seconds
(without alignments)
17784.976 Million cell updates/sec

Title: US-09-809-920-3

Perfect score: 6172
Sequence: 1 GCGGGCTCCTGAGCTGAA.....ATMAATGAGACTTTAACTC 6172

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 segs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seg length: 0
Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	6172	100.0	6172	9	AB007042	Homo sapi
2	6172	100.0	6172	9	AF083551	Homo sapi
3	6163.8	99.9	6189	9	AB011091	Homo sapi
4	6124.8	99.2	6147	9	AF001690	Homo sapi
5	5834.6	94.5	5854	9	AF029231	Homo sapi
6	3217	52.1	3241	9	BC006363	Homo sapi
7	3035	49.2	167343	6	AX332231	Sequence
8	3035	49.2	167343	6	AX335067	Sequence
9	3035	48.9	168194	2	HS096629	Human chrom
10	3018.4	41.6	199912	2	AC040975	Homo sapi
11	2567.4	39.9	3479	10	AC012411	Homo sapi
12	2459.6	38.0	3198	10	AF083550	Mus muscu
13	2347.4	36.0	3198	10	AB033367	Rattus no
14	1077.8	17.5	236447	2	AC087099	Mus muscu
15	621	10.1	64626	2	AC011641	Homo sapi
16	596.4	9.7	598	6	AX062596	Sequence
17	545.6	8.8	3622	3	AF132161	Drosophill
18	545.6	8.8	106659	2	AC020331	Drosophill
19	545.6	8.8	161172	3	AC007839	Drosophill
20	545.6	8.8	287859	3	AE003797	Drosophill
21	521.8	8.5	64626	2	AC011641	Homo sapi
22	396	6.4	396	9	HS076188	Homo sapi
23	256.2	4.2	2530	3	CE094835	Caenorhabd
24	256.2	4.2	2616	3	CE094835	Caenorhabd
25	244.6	4.0	23164	2	AC108250	Rattus no
26	224.4	3.6	2860	10	MM067837	Mus muscu
27	224.4	3.6	2862	10	BC006597	Mus muscu
28	224.4	3.6	2864	10	MM072141	Mus muscu
29	218.8	3.5	2884	4	AF089748	Bos tauru
30	200.6	3.3	3069	9	BC010058	Homo sapi
31	200.6	3.3	3175	9	HS062740	Homo sapi
32	200.6	3.3	3320	9	HS064511	Human herd
33	199	3.2	3003	6	AR168099	Sequence
34	199	3.2	3003	9	HS072263	Human multi
35	181.2	2.9	3077	10	AF252858	Cricetulu
36	174.6	2.8	2562	10	BC004741	Mus muscu
37	174.6	2.8	2708	10	MM078539	Mus muscu
38	169.4	2.7	2769	9	BC001174	Homo sapi
39	169.4	2.7	3183	9	S79639	EXT1-pulati
40	166.2	2.7	2709	3	AF145598	Drosophill
41	166.2	2.7	3642	3	AF083889	Drosophill
42	160.2	2.6	2965	6	AX306135	Sequence
43	160.2	2.6	2965	10	MMEX11	Mus muscu
44	151	2.4	2867	5	AF319538	Xenopus 1
45	140.6	2.3	19442	3	CEK01G5	Caenorhabd

ALIGNMENTS

RESULT 1
LOCUS AB007042 6172 bp mRNA linear PRI 13-FEB-1999
DEFINITION Homo sapiens EXTR1 mRNA, complete cds.
ACCESSION AB007042
VERSION AB007042.1 GI:2723390
KEYWORDS EXTR1.
SOURCE Homo sapiens testis and total fetus cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 6172)
AUTHORS Saito,T.
TITLE Direct Submision
JOURNAL Submitted (06-SEP-1997) 'Toshiyuki Saito, National Institute of
Radiological Sciences, Genome Research Group; Anagawa 4-9-1, Inage,
Chiba 263-8555, Japan (E-mail: t-saito@nirs.go.jp,
Tel:81-43-206-3135, Fax:81-43-251-9818)
REFERENCE 2 (sites)
AUTHORS Saito,T., Seki,N., Yamauchi,M., Tsuji,S., Hayashi,A., Kozuma,S. and
Horii,T.
TITLE Structure, chromosome location, and expression profile of EXTR1
and EXTR2, new members of the multiple exostoses gene family

JOURNAL Biochem. Res. Commun. 243 (1), 61-66 (1998)
MEDLINE 98139867
COMMENT updated (19-Sep-1997).
FEATURES Location/Qualifiers
source 1. 6172
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="8p"
/tissue_type="testis and total fetus"
594..3353
/gene="EXTR1"
594..3353
/gene="EXTR1"
/codon_start=1
/product="EXTR1"
/protein_id="BAA24080.1"
/db_xref="GI:2723391"
/translation="MTGYTMLRNGAGNGGOTCMLRWSNRILRWLSFTFLVILVFP
LIAHYITLLEADBAKGRIFGRVGNELCEVHLDLCRISESSEILOEAKROE
LNSEIATLNLKIEACKRSIENAKODLQIKNYISOTESHYKELMONGPKISLIRL
PEKDDAGLPKPRATGCRGHCNCFDYSRCLPSGPPVYVDSQPFYSGLDPLVQAF
QATARANYTEMNDIACLYIVLIVSEMPVYRPAELKSLTSLPHRTGHNHVI
NLSKSDQNLINLVSTRAWAOSTFTYVYRQFDLVVSPLVHAMEPNMEIPQ
VPAKRYLFTQGEKIESLRSLQEARFEEEMEDPADYDRIATLKAQVDSKLD
QVLEFETCKNPKSLPTHEMALCGEREDRLKISTFALLITPDPLVSSGCAT
LFEALVGAIVVGLGEVOLPYODMLQNEALVPIKREVRVHEHLISDSOLLAM
RROGFLEMETEYSTADSIENYVILMIRRIIPAPIREPAAEATPHSGRAAGTDPN
MADGDDIGVETPEPYASPRYLKNTFLTYTDYRSMNCAPFHLPHFDFVLP
SEAFPLSGSTFRPIGGAGSGSGKEFCALGQVPRQETVVMILYEVEVLMNSLER
LNGLEPLKVVVYVNSKPLPSIEDLMPDIPVIMVARTKNSLNNRFLPMNEIEAI
LSIDDAHLRHDEIMFGFRVREARDRIVGPGRHADIHQSLVNSNSCELSW
LTGAAFHKKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFIIVSHITRPPKIVSRMT
FRGCGCPQALSHDSDSHFERHKCINFEKYYGVMPILYTOFRVDSVLEFKTLPHDKTK
CFKFI"

BASE COUNT 1280 a 1681 c 1692 g 1519 t
ORIGIN

Query Match 100.0%; Score 6172; DB 9; Length 6172;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGGGTCCTGAGCTGGAAGCCGAGAGCAAGCCCTGGAGTTCACCTTTCAAGAGT 60
DB 1 GCGGGGTCCTGAGCTGGAAGCCGAGAGCAAGCCCTGGAGTTCACCTTTCAAGAGT 60
QY 61 CGTGGCTGAGAGTGTAAATGCTACACAGTCAAGAGAGAGAGGCTCTGAACAACATGCG 120
DB 61 CGTGGCTGAGAGTGTAAATGCTACACAGTCAAGAGAGAGAGGCTCTGAACAACATGCG 120
QY 121 CTGATTGTGGCAAGGCATCATTAAGAGCTGGGCAATTATTCTGTCCTTACCTATTACT 180
DB 121 CTGATTGTGGCAAGGCATCATTAAGAGCTGGGCAATTATTCTGTCCTTACCTATTACT 180
QY 181 GTATAACTGTGAATAGACACTATGATTTGTGTGACGCAAAACCAAGAAAGAGC 240
DB 181 GTATAACTGTGAATAGACACTATGATTTGTGTGACGCAAAACCAAGAAAGAGC 240
QY 241 TATGGCAATTTGAAGAAAGTGTGTGATTTCCAGGCTGTTTCTCGGTTTCATCATCAGG 300
DB 241 TATGGCAATTTGAAGAAAGTGTGTGATTTCCAGGCTGTTTCTCGGTTTCATCATCAGG 300
QY 301 TACCTCTCCCTTCATCTCAGCAAGAAATGGGCACTTTTATGCTTTGATTAAGATTAA 360
DB 301 TACCTCTCCCTTCATCTCAGCAAGAAATGGGCACTTTTATGCTTTGATTAAGATTAA 360
QY 361 GGAACATGTTCTTTGGTCAACAGCCAGAACTTAAATCTGCGGAATGGGGCAGAGACCA 420
DB 361 GGAACATGTTCTTTGGTCAACAGCCAGAACTTAAATCTGCGGAATGGGGCAGAGACCA 420
QY 421 TTTTCAGCTGCAGCTGAGAGAAATGAATGTTCAATTTTATTTGGTCTTCTGGGAGC 480
DB 421 TTTTCAGCTGCAGCTGAGAGAAATGAATGTTCAATTTTATTTGGTCTTCTGGGAGC 480

DB 421 TTTTCAGCTGCAGCTGAGAGAAATGAATGTTCAATTTTATTTGGTCTTCTGGGAGC 480
QY 481 ACACATACTCTTCTGGAACGCTGTCACTGAACAGAGATGTTTGTGAATAAGCAACC 540
DB 481 ACACATACTCTTCTGGAACGCTGTCACTGAACAGAGATGTTTGTGAATAAGCAACC 540
QY 541 ATGGTATGCGAGTGAACCCGAGCTGATCTGGGGGGCAGGCTTCAGAGACTATACAG 600
DB 541 ATGGTATGCGAGTGAACCCGAGCTGATCTGGGGGGCAGGCTTCAGAGACTATACAG 600
QY 601 GCTATACATGCTGCGGAATGAGGGGCGGGGAAAGAGAGTGCAGCTGATCTGCGCT 660
DB 601 GCTATACATGCTGCGGAATGAGGGGCGGGGAAAGAGAGTGCAGCTGATCTGCGCT 660
QY 661 GGTCCACCGCATCCGCTCTACGCTGAGCTGACCTTCAGCTCTTGTATCTGCTGCT 720
DB 661 GGTCCACCGCATCCGCTCTACGCTGAGCTGACCTTCAGCTCTTGTATCTGCTGCT 720
QY 721 TCCCGCTCATGCGCCCTATTACCTCACCACCTTGTGATGAGCTGATGAGGAGCAGC 780
DB 721 TCCCGCTCATGCGCCCTATTACCTCACCACCTTGTGATGAGCTGATGAGGAGCAGC 780
QY 781 GGAATTTTGTGTCGCCGGTGGGGAAAGAGCTGTGCAAGGACGCTGATCTGT 840
DB 781 GGAATTTTGTGTCGCCGGTGGGGAAAGAGCTGTGCAAGGACGCTGATCTGT 840
QY 841 GCCGCATCCGGAGATCGGTGAGTGAAGAGCTCTCTGAGCTGAGGCGCAAGCCCAAGC 900
DB 841 GCCGCATCCGGAGATCGGTGAGTGAAGAGCTCTCTGAGCTGAGGCGCAAGCCCAAGC 900
QY 901 TGAACAGGAGATTCGCAAGCTGAAATGTAATGAAAGCCCTGTAAAGAGACATTGAGA 960
DB 901 TGAACAGGAGATTCGCAAGCTGAAATGTAATGAAAGCCCTGTAAAGAGACATTGAGA 960
QY 961 ACGCAAGAGAGACCTGCTCCAGCTCAAGAAATGTATCAGCAGCAGCCGACATTCCTCA 1020
DB 961 ACGCAAGAGAGACCTGCTCCAGCTCAAGAAATGTATCAGCAGCAGCCGACATTCCTCA 1020
QY 1021 AGGAGCTATGCGCCCAAGAACAGCCCAAGCTGCTCCCTCCATCCGATGCTCCCAAGA 1080
DB 1021 AGGAGCTATGCGCCCAAGAACAGCCCAAGCTGCTCCCTCCATCCGATGCTCCCAAGA 1080
QY 1081 AGGAGATGCGCGGCTCTCCCTCCCGGAAGGCGACCTGCGGCTACACAACCTCT 1140
DB 1081 AGGAGATGCGCGGCTCTCCCTCCCGGAAGGCGACCTGCGGCTACACAACCTCT 1140
QY 1141 TTGATTTATTCCTGTTGCCCTCTCACTCTGCTTCCGCTCTACGCTATGACAGTGC 1200
DB 1141 TTGATTTATTCCTGTTGCCCTCTCACTCTGCTTCCGCTCTACGCTATGACAGTGC 1200
QY 1201 AGTTGTCTTTGGCAGCTACCTGATCCCTTGTGTCAAGCAGGCTTTTCAGGCGACAGC 1260
DB 1201 AGTTGTCTTTGGCAGCTACCTGATCCCTTGTGTCAAGCAGGCTTTTCAGGCGACAGC 1260
QY 1261 GAGCTAAGCTTATGTTACAAATAATGACATGCGCTGCTTACGTATACATAGTG 1320
DB 1261 GAGCTAAGCTTATGTTACAAATAATGACATGCGCTGCTTACGTATACATAGTG 1320
QY 1321 GAGAGATGCAAGGAGCCGCTGCTGCGGCTGTGAGAGCTGAGAGAGAGTGTATTC 1380
DB 1321 GAGAGATGCAAGGAGCCGCTGCTGCGGCTGTGAGAGCTGAGAGAGAGTGTATTC 1380
QY 1381 TGCCACACTGGCGGAGCGATGACACAACATGTCATCATCAATCTGTCAAGTAACTAG 1440
DB 1381 TGCCACACTGGCGGAGCGATGACACAACATGTCATCATCAATCTGTCAAGTAACTAG 1440
QY 1441 ATACACAGAACCTTCTATTAAGCTCACTGCTGCGCTGCTGAGAGAGAGTGTATTC 1500
DB 1441 ATACACAGAACCTTCTATTAAGCTCACTGCTGCGCTGCTGAGAGAGAGTGTATTC 1500
QY 1501 TCTACACTGTCAGTACAGAGCTGCGCTTACCTTGTGCTGATACAGCAGCTGCTCAGTGC 1560
DB 1501 TCTACACTGTCAGTACAGAGCTGCGCTTACCTTGTGCTGATACAGCAGCTGCTCAGTGC 1560

Qy	1561	TGTCGTAGGCCAACTTCATGAAATCCACACAGGTGCCGGTGAAGCGGAATATCTCT	1620
Db	1561	TT	1620
Qy	1561	TGTCGTAGGCCAACTTCATGAAATCCACACAGGTGCCGGTGAAGCGGAATATCTCT	1620
Qy	1621	TCACGCTCCAGGGCGAAGATTGAGTCTGTAGGTCGTAGCCCTTCAGAGAGCCGGCTCT	1680
Db	1621	TT	1680
Qy	1681	TCGAAAGAGAAATGAGAGGGCGACCTCCCGCGACTACGATGACCGGATCATTTGCCACCC	1740
Db	1681	TCGAAAGAGAAATGAGAGGGCGACCTCCCGCGACTACGATGACCGGATCATTTGCCACCC	1740
Qy	1741	TCGAAAGCGGTGCAGAGCAGCAGAGCTGGATTCAGGTCTGGTGGAAATTCACCTGCAGAAAACC	1800
Db	1741	TCGAAAGCGGTGCAGAGCAGCAGAGCTGGATTCAGGTCTGGTGGAAATTCACCTGCAGAAAACC	1800
Qy	1801	AGCCCAAAACCCAGCCCTGCCGACTGATGGGCACTGTGTGAGAGCGGAGAGCCGCTTGG	1860
Db	1801	AGCCCAAAACCCAGCCCTGCCGACTGATGGGCACTGTGTGAGAGCGGAGAGCCGCTTGG	1860
Qy	1861	AAATGCTGAAGCTCTCCACCTTCGCCCTCATCATTAACCCCGGGGAGCCCTCGCTTGGTTA	1920
Db	1861	AAATGCTGAAGCTCTCCACCTTCGCCCTCATCATTAACCCCGGGGAGCCCTCGCTTGGTTA	1920
Qy	1921	TTTTCTCTGGGTGTGCAACAAGCGCTCTTGAAAGCCCTGGAAGTTCGGTCCGCTGCCGTGG	1980
Db	1921	TTTTCTCTGGGTGTGCAACAAGCGCTCTTGAAAGCCCTGGAAGTTCGGTCCGCTGCCGTGG	1980
Qy	1981	TGCTGGGGGAGCAGTCGCAAGCTTCCCTACAGAGACATGCTGCAAGTGGAAAGAGCGGCC	2040
Db	1981	TGCTGGGGGAGCAGTCGCAAGCTTCCCTACAGAGACATGCTGCAAGTGGAAAGAGCGGCC	2040
Qy	2041	TGGTGGTGCAGAAAGCCTCGTGTACCGAGGTTCAATTCCTGCTCAGAAAGCCTTCGCGATA	2100
Db	2041	TGGTGGTGCAGAAAGCCTCGTGTACCGAGGTTCAATTCCTGCTCAGAAAGCCTTCGCGATA	2100
Qy	2101	GTGACCTCTGGGTATGAGAGCGGCAAGCGGCTTCTCTGGGAGACTTACTTCTCCACTG	2160
Db	2101	GTGACCTCTGGGTATGAGAGCGGCAAGCGGCTTCTCTGGGAGACTTACTTCTCCACTG	2160
Qy	2161	GTGACAGTATTTTTTAATTAACCGTCTGCTGGCTATGATTAAGAGATCGCATCCAGATCCACCG	2220
Db	2161	GTGACAGTATTTTTTAATTAACCGTCTGCTGGCTATGATTAAGAGATCGCATCCAGATCCACCG	2220
Qy	2221	CTCCCATCCGGGGAAGAGCGGCGAGCTGAGATCCGCCACCGTTAGGCAAGCGAGCGCTGGAA	2280
Db	2221	CTCCCATCCGGGGAAGAGCGGCGAGCTGAGATCCGCCACCGTTAGGCAAGCGAGCGCTGGAA	2280
Qy	2281	CTGACCCCAACATGCGTGTGACACAGCGGAGCTGGACCTGGGGCCAGTGTGAAACGAGAGCCG	2340
Db	2281	CTGACCCCAACATGCGTGTGACACAGCGGAGCTGGACCTGGGGCCAGTGTGAAACGAGAGCCG	2340
Qy	2341	CGTAGCGCTACACCAGATACCTCCGCAATTTCACTGTGACTGTACACTGACTTTTACC	2400
Db	2341	CGTAGCGCTACACCAGATACCTCCGCAATTTCACTGTGACTGTACACTGACTTTTACC	2400
Qy	2401	GGTGAAGTGTGCTCCAGGCGCTTTCCATCTTTTCCGCCACACCTCCCTTTGACCCCTGTGT	2460
Db	2401	GGTGAAGTGTGCTCCAGGCGCTTTCCATCTTTTCCGCCACACCTCCCTTTGACCCCTGTGT	2460
Qy	2461	TGCGCTCAGAGGCCAAATTCCTGGGCTCAGAGACTGGCTTGGGCTATTTGGTGTGGAG	2520
Db	2461	TGCGCTCAGAGGCCAAATTCCTGGGCTCAGAGACTGGCTTGGGCTATTTGGTGTGGAG	2520
Qy	2521	CTGGGGGTTCTGGCAAGGAATTTTCAGGCAAGCGCTTGGAGCAATTTTCCCGAGAGAGAGT	2580
Db	2521	CTGGGGGTTCTGGCAAGGAATTTTCAGGCAAGCGCTTGGAGCAATTTTCCCGAGAGAGAGT	2580
Qy	2581	TCACGGTGTGATGTTGACTTATGACGCGGAGAGAGAGTGTATTAAGACTCTTTAGAGAGGC	2640
Db	2581	TCACGGTGTGATGTTGACTTATGACGCGGAGAGAGAGTGTATTAAGACTCTTTAGAGAGGC	2640

QY	2641	TGATGGCCCTCCCTTACCTTAACAAGGTGTTGGTGGTGGAAATTCCTCCCAAGCTGCAT	2700
Db	2641	TGAATGGCCCTCCCTTACCTTAACAAGGTGTTGGTGGTGGAAATTCCTCCCAAGCTGCAT	2700
QY	2701	CAGAGGACCTTCGTGGCCCTGACATTGGGGTTCCCACTCATGGGNGCCGTACTGGAAGA	2760
Db	2701	CAGAGGACCTTCGTGGCCCTGACATTGGGGTTCCCACTCATGGTGGTCCGTACTGGAAGA	2760
QY	2761	ACAGTTTGAACAACCGATTCTTACCCCTGGAAATGAAATTGAGACAGAGGCCATCTGTCCA	2820
Db	2761	ACAGTTTGAACAACCGATTCTTACCCCTGGAAATGAAATTGAGACAGAGGCCATCTGTCCA	2820
QY	2821	TTGATGACATGCTCACTCCGCCATGACGAATCATGTTTGGGTCCGGGTGTGGAAG	2880
Db	2821	TTGATGACATGCTCACTCCGCCATGACGAATCATGTTTGGGTCCGGGTGTGGAAG	2880
QY	2881	AAGCTGGGAGCCGATGCTGGGGCTCCCTGGCCGTTACCACGCATGAGGACATCCCCATTC	2940
Db	2881	AAGCTGGGAGCCGATGCTGGGGCTCCCTGGCCGTTACCACGCATGAGGACATCCCCATTC	2940
QY	2941	AGTCTGGCTCTTACAACTCCAACTACTCTGTGAGCTGTCCATGGTGTGACAGGTGCTG	3000
Db	2941	AGTCTGGCTCTTACAACTCCAACTACTCTGTGAGCTGTCCATGGTGTGACAGGTGCTG	3000
QY	3001	CCCTCTTTCACAAATATTATAGCCTACCTGTATTCTTATGTATGATGCCACAGCCATCCGG	3060
Db	3001	CCCTCTTTCACAAATATTATAGCCTACCTGTATTCTTATGTATGATGCCACAGCCATCCGG	3060
QY	3061	ACATGTGGAGTATACATCACTCACTGTGAGGACATCCCATGAAATCCCTGTGCTCCACA	3120
Db	3061	ACATGTGGAGTATACATCACTCACTGTGAGGACATCCCATGAAATCCCTGTGCTCCACA	3120
QY	3121	TCACCTGGGAAGCCCCCATCAAGGTGACCTCAACAGGTGACATTCGATCCGAGATGCC	3180
Db	3121	TCACCTGGGAAGCCCCCATCAAGGTGACCTCAACAGGTGACATTCGATCCGAGATGCC	3180
QY	3181	CTCAGGCCCTGTCTCATGATGACTCCCACTTCACAGAGCGGACAAATGCATCAACTTCT	3240
Db	3181	CTCAGGCCCTGTCTCATGATGACTCCCACTTCACAGAGCGGACAAATGCATCAACTTCT	3240
QY	3241	TCGTGAAGGTGTAGGGCTAATATGCCCTCCTGTACACGCAAGTTCAGGGTGGATTTCTGTC	3300
Db	3241	TCGTGAAGGTGTAGGGCTAATATGCCCTCCTGTACACGCAAGTTCAGGGTGGATTTCTGTC	3300
QY	3301	TCCTTCAAAGACAGCGCTCCCATGACAAAGACCAAGTTCATCTAAGTGGGGAGAC	3360
Db	3301	TCCTTCAAAGACAGCGCTCCCATGACAAAGACCAAGTTCATCTAAGTGGGGAGAC	3360
QY	3361	GCACGCTCTGGGGAAGAAGATGAGACAGAGGAGGAAGATGGTCCCAAGTTCCTTAGCA	3420
Db	3361	GCACGCTCTGGGGAAGAAGATGAGACAGAGGAGGAAGATGGTCCCAAGTTCCTTAGCA	3420
QY	3421	TTGCAGGACCTTGGGCACTCTGCTGGTGGTGGGCCCAAGCCCTGTCTGGAAGGGGAC	3480
Db	3421	TTGCAGGACCTTGGGCACTCTGCTGGTGGTGGGCCCAAGCCCTGTCTGGAAGGGGAC	3480
QY	3481	CAGAGAGAGTGGAAAGGAAACCGCTGTATTCTTAAATCAAGCCACATCGGGCCTTGAG	3540
Db	3481	CAGAGAGAGTGGAAAGGAAACCGCTGTATTCTTAAATCAAGCCACATCGGGCCTTGAG	3540
QY	3541	CCCTGGGGGAGTCCCGGGGGTTCCCAACAGAGGACAGTGAATGATTTCACTGAGG	3600
Db	3541	CCCTGGGGGAGTCCCGGGGGTTCCCAACAGAGGACAGTGAATGATTTCACTGAGG	3600
QY	3601	ACTGTGGGACTCTGCAGAGTCACTCAACCGTTCCTACGCCCAAGACAGCTGTCTGTG	3660
Db	3601	ACTGTGGGACTCTGCAGAGTCACTCAACCGTTCCTACGCCCAAGACAGCTGTCTGTG	3660
QY	3661	GTTTTTACATTCAATAACAACATATTATGATTTATTTAAAAAGGAAAGTTTCAGATTTACC	3720
Db	3661	GTTTTTACATTCAATAACAACATATTATGATTTATTTAAAAAGGAAAGTTTCAGATTTACC	3720
QY	3721	ATTCGAAGCTATTATTATATTATGTTGGTATATATATTAATATATGACACACTGTGCATGA	3780

D	b		3721	ATTCAAGCGCTTATTTATATATATATGTCGTGAATAAATAACATGCACACTTGATACA	3780
O	y		3781	TATATATTTTTGGCTGGGGAGATGTAGTTTTGCCCTTCTTAAGGAGAGACGGGACGC	3840
D	b		3781	TATATATTTTTGGCTGGGGAGATGTAGTTTTGCCCTTCTTAAGGAGAGACGGGACGC	3840
O	y		3841	TCCCTTGTCCTATTTCTGCGGAGATGGTTCGGCCTTGTCACATGCTTATCCCTTA	3900
D	b		3841	TCCCTTGTCCTATTTCTGCGGAGATGGTTCGGCCTTGTCACATGCTTATCCCTTA	3900
O	y		3901	AAGATCATCTCCCATCTCCACCAGGGCATCTGTGTGTGAGAACACAGAAGGATGAAT	3960
D	b		3901	AAGATCATCTCCCATCTCCACCAGGGCATCTGTGTGTGAGAACACAGAAGGATGAAT	3960
O	y		3961	TGGCCCTCTGCGGGCCGTGACAAAGGTCTCTCCCTTACCTTCTGTGGCAGTACCAA	4020
D	b		3961	TGGCCCTCTGCGGGCCGTGACAAAGGTCTCTCCCTTACCTTCTGTGGCAGTACCAA	4020
O	y		4021	CCTGTAACTCATTTCTTCTCCAGTAAATCCCTGGAGGCCCTGACCTCGTGGGCTGT	4080
D	b		4021	CCTGTAACTCATTTCTTCTCCAGTAAATCCCTGGAGGCCCTGACCTCGTGGGCTGT	4080
O	y		4081	TCAGCTTCCTCTCTCTGGGGCACAGATTTTTTAGAGATTATCTTtagccagccttgcc	4140
D	b		4081	TCAGCTTCCTCTCTCTGGGGCACAGATTTTTTAGAGATTATCTTtagccagccttgcc	4140
O	y		4141	TCCGTAATATCCCTGCTCTCCCATTTCTCTGTTTGAAGAGATGAGGAAGAAG	4200
D	b		4141	TCCGTAATATCCCTGCTCTCCCATTTCTCTGTTTGAAGAGATGAGGAAGAAG	4200
O	y		4201	AGTAGAAGAAATAAGGGGCTGAAGAGCCACATCCACAGATGGCTTTTCTATCCTGCTT	4260
D	b		4201	AGTAGAAGAAATAAGGGGCTGAAGAGCCACATCCACAGATGGCTTTTCTATCCTGCTT	4260
O	y		4261	CTGTGTAACACACATGTCGTGTGGGCTCCAGGGCTTCTGAAGGCTCTTTTGGAATTGG	4320
D	b		4261	CTGTGTAACACACATGTCGTGTGGGCTCCAGGGCTTCTGAAGGCTCTTTTGGAATTGG	4320
O	y		4321	ACAGAGATCAGCAGCGTCACATCTGCTGTGTGAAGTGGTTTGACAGTCACGCTCC	4380
D	b		4321	ACAGAGATCAGCAGCGTCACATCTGCTGTGTGAAGTGGTTTGACAGTCACGCTCC	4380
O	y		4381	TCGCCCTAGTGTAGACAGCCAGTCGTCTTGAGAGAACCCCGCGCTGGCCGGGAAGT	4440
D	b		4381	TCGCCCTAGTGTAGACAGCCAGTCGTCTTGAGAGAACCCCGCGCTGGCCGGGAAGT	4440
O	y		4441	TTTTACAGCAAGGCGCTGCTGGGATTAATCTGTGGAAATTCACCTTCCCCCGCT	4500
D	b		4441	TTTTACAGCAAGGCGCTGCTGGGATTAATCTGTGGAAATTCACCTTCCCCCGCT	4500
O	y		4501	CTGTCTGGAGCCCATCTGTGTATCTGTGTGTTTTTGACCCCTTAATGTACGCTGGCT	4560
D	b		4501	CTGTCTGGAGCCCATCTGTGTATCTGTGTGTTTTTGACCCCTTAATGTACGCTGGCT	4560
O	y		4561	GTAGAGCTCCCGAGGTTTGGATATGCTAGAAACATGTGAAGGCTGTGATTTGCTGTGA	4620
D	b		4561	GTAGAGCTCCCGAGGTTTGGATATGCTAGAAACATGTGAAGGCTGTGATTTGCTGTGA	4620
O	y		4621	AGCTACATTCACGCTTGGAATCTPAACGGGGATTCACACCCCGAGTTACACATTCCACT	4680
D	b		4621	AGCTACATTCACGCTTGGAATCTPAACGGGGATTCACACCCCGAGTTACACATTCCACT	4680
O	y		4681	CCCTCTTAGAGATTCGTTCCTGGGCTGAACATGAATAAGCTAATTTTTGGGTACG	4740
D	b		4681	CCCTCTTAGAGATTCGTTCCTGGGCTGAACATGAATAAGCTAATTTTTGGGTACG	4740
O	y		4741	GTGGCAGTAGGGAACCTGAGAGGGTGTGATGGTCATTTGTCAAGGATTTAGCCCATGAC	4800
D	b		4741	GTGGCAGTAGGGAACCTGAGAGGGTGTGATGGTCATTTGTCAAGGATTTAGCCCATGAC	4800
O	y		4801	GTGTTTCTGAACCTACTTCTCTGGAAGTGAAGTGTGAGGTTTCTAGCACTG	4860
D	b		4801	GTGTTTCTGAACCTACTTCTCTGGAAGTGAAGTGTGAGGTTTCTAGCACTG	4860

[illegible]

RESULT 2
AF083551
LOCUS Homo sapiens exostosin (EXTL3) mRNA, complete cds.
DEFINITION AF083551
ACCESSION AF083551
VERSION AF083551.1 GI:5326770
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 6172)
AUTHORS Sato, T.
TITLE The human EXTL3/EXTL3 gene at Bp11-p12, the third breast cancer susceptibility gene locus, is not mutated in breast and various cancers
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 6172)
AUTHORS Sato, T.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1998) Otolaryngology/Pathology, Columbia University, 630 West 168th St., P&S 11-451, New York, NY 10032, USA
FEATURES
SOURCE
1..6172
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="Bp11-p12"
1..6172
/gene="EXTL3"
/note="EXTL3; EXTL3; candidate gene for breast cancer susceptibility BRCA3"
594...3353
/gene="EXTL3"
/codon_start=1
/product="exostosin"
/db_xref="GI:5326771"
/db_xref="AA042041.1"
Translation: MTCYMLRNNGNGSGOTCMILMNSNRIILTWLSFTLVILVEEP
LIAHYVLTLDDEADGRIEFGVNELEKIVLDLCIRVSESELLEKROE
LNSFIKLTKEACKKSEINAKODLILNVISOTHSKELNMLNPLPILRL
PERDAGLPPKATRGCRNLNCSKPLTSGPRVYVSDOPRESYIDPLVKAF
QATRAVYVTEADINACLVILGEMQEVNLPALAEQLSLPHWRDGHVIT
NLSKSDTONLNVSTGRAMVAOSTEYVQYRGFDLVVSPVHAASENMEIPQ
VPAKRYLFTFOGKIESLSLOEARSFEEMGDPADYDRIATLKAVIDSKID
QVLEVFKNQPKPSLPTEWALGSEDRDLKLSFALLITPGDRLVYSSGAR
LFEALVGAVPVVGEOVLGYDMLQWNAALVPAKRYTEVAFILRLSDSLAM
RRCGRLEMTVEFSTADISNTVLAITRTQIPAAPIREAAAEIPRISKAQTDN
MADNGDLDLGPETEPPYASPRILRNFTLVITDYSNWKACPGFHLPTPTDPLV
SEAKFLSGTGFRTIGGAGSGKEFOALAGVPRQFTVMTYREVENLANSLE
LNGLPYLNKVVVWNSPKLPSSEDLMPDIDVPIVNVTEKNSLNNRPLPNNIETEA
LSIDDAHLRHDEIMFGFVYMRARDRIVGFGRYHAMDIPHOSMLYNSYSCLEW
LTGAAPFKYVAYLYSVMPQAIRDWDVEYINCEDIAMNLFVITTRKPIKYSMT
FRCGCCPOLASHSDSHFHRKCNIFVVKYGVYPLLYTQPRVDSVLFKRLPHDKK
CFKFT

BASE COUNT 1280 a 1681 c 1692 g 1519 t
ORIGIN
Query Match 100.0%; Score 6172; DB 9; Length 6172;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCGGGTCCCTGAGCTGGAAAGCCGAGAGAGCAAGCCCTGAGAGTTCACCTTTCAAGAGT 60
Db 1 GCGGGTCCCTGAGCTGGAAAGCCGAGAGAGCAAGCCCTGAGAGTTCACCTTTCAAGAGT 60
QY 61 CGTGTCTGAGCTGTAATGCTACACAGTCAGAGAGAGAGAGAGAGTCTCTGAACACATGCG 120
Db 61 CGTGTCTGAGCTGTAATGCTACACAGTCAGAGAGAGAGAGAGAGTCTCTGAACACATGCG 120
QY 121 CTGATGTGTGGCAAGCATATGAAGCGGATTTATTTCTGTTCACTTAATTAAT 180
Db 121 CTGATGTGTGGCAAGCATATGAAGCGGATTTATTTCTGTTCACTTAATTAAT 180
QY 181 GTATTAAGTGAATAGACATATGATATTTGTTGTCAGCAAAACAGAAAGAGC 240
Db 181 GTATTAAGTGAATAGACATATGATATTTGTTGTCAGCAAAACAGAAAGAGC 240
QY 241 TATGGCATTTGAAAAAGCTGTGATTCAGAGGATTTTCTGGGTTTCATCATCAGG 300
Db 241 TATGGCATTTGAAAAAGCTGTGATTCAGAGGATTTTCTGGGTTTCATCATCAGG 300
QY 301 TACCTCCCTCTTCATCTCAGCAAGAAATGTGCACTTTATGCTTTGATTAAGATTAA 360
Db 301 TACCTCCCTCTTCATCTCAGCAAGAAATGTGCACTTTATGCTTTGATTAAGATTAA 360
QY 361 GGACATGTTCTTGTGTCACAGCAGCAAGCTTAATGCTGTAATAGGTCAGAGACA 420
Db 361 GGACATGTTCTTGTGTCACAGCAGCAAGCTTAATGCTGTAATAGGTCAGAGACA 420
QY 421 TTTAGCTGACGCTGAGAGAAATGAATGTCATTTATTTGTTGCTGTGGGAGC 480
Db 421 TTTAGCTGACGCTGAGAGAAATGAATGTCATTTATTTGTTGCTGTGGGAGC 480
QY 481 ACATTAAGTCTTGTGGAAGCTGTCAGTAAGAAAGAGAGAGTGTGTTGGAATACCAACC 540
Db 481 ACATTAAGTCTTGTGGAAGCTGTCAGTAAGAAAGAGAGAGTGTGTTGGAATACCAACC 540
QY 541 ATGGTATAGGAGTGAACCCGACGTCGATCTGGGGGAGAGCTGACAGAGACTCATGACAG 600
Db 541 ATGGTATAGGAGTGAACCCGACGTCGATCTGGGGGAGAGCTGACAGAGACTCATGACAG 600
QY 601 GCTATACCATGCTCGGAGATGGGGGCGGGGAGACGAGTCAAGACCTGCATCGTGGCGT 660
Db 601 GCTATACCATGCTCGGAGATGGGGGCGGGGAGACGAGTCAAGACCTGCATCGTGGCGT 660
QY 661 GGTCACACCGCATCGGCTCGTCAAGGCTCAGCTTCAGCTTGTATCTGATCTTCT 720
Db 661 GGTCACACCGCATCGGCTCGTCAAGGCTCAGCTTCAGCTTGTATCTGATCTTCT 720
QY 721 TCCGCTCATATGCGCCACTATATACCTCAACACTGATGAGAGCTGATGAGGACAGCAAGC 780
Db 721 TCCGCTCATATGCGCCACTATATACCTCAACACTGATGAGAGCTGATGAGGACAGCAAGC 780
QY 781 GGATTTTGTGCTCCCGGGTGGGAGACGAGTGTGCGAGAGTGAACGACGTGCTGATCTGT 840
Db 781 GGATTTTGTGCTCCCGGGTGGGAGACGAGTGTGCGAGAGTGAACGACGTGCTGATCTGT 840
QY 841 GCGCATCCCGGAGATCGGTGATGAAGAGCTCGAGAGCTGAGAGGCAAGGCGCAAGAGC 900
Db 841 GCGCATCCCGGAGATCGGTGATGAAGAGCTCGAGAGCTGAGAGGCAAGGCGCAAGAGC 900
QY 901 TGAACAGAGAGATGCGCAAGCTGATGAGATGAGAGCTGTAAGAGAGATTTAGAG 960
Db 901 TGAACAGAGAGATGCGCAAGCTGATGAGATGAGAGCTGTAAGAGAGATTTAGAG 960
QY 961 AGCGCAAGCAGAGACCTGCTCCAGCTCAAGAAATGTCATCAGCAGAGCAGCATTTCTTACA 1020

```
|||||
Db 961 ACCCAAGCAGACCTCTCAGCTCAAGATGTCATCAGCAGACGACGATTTCTACA 1020
QY 1021 AGGAGCTATGAGCCAGAAACACAGCTGTCTCCATCCGACTCTCCACAGA 1080
Db 1021 AGGAGCTATGAGCCAGAAACACAGCTGTCTCCATCCGACTCTCCACAGA 1080
QY 1081 AGGAGCAGCGGCGCTCCCTCCCGAAGGCACTGGGGCTGCGGCTACAACTGCT 1140
Db 1081 AGGAGCAGCGGCGCTCCCTCCCGAAGGCACTGGGGCTGCGGCTACAACTGCT 1140
QY 1141 TTGATTAATCTGCTGCTCCCTCTCAGCTGTGCTTCCGGTCTACGTCTATGACAGTAC 1200
Db 1141 TTGATTAATCTGCTGCTCCCTCTCAGCTGTGCTTCCGGTCTACGTCTATGACAGTAC 1200
QY 1201 AGTTGTCTTTGGAGCTACCTGATCCCTTGGTCAAGAGCTTTTACAGGACAGCAC 1260
Db 1201 AGTTGTCTTTGGAGCTACCTGATCCCTTGGTCAAGAGCTTTTACAGGACAGCAC 1260
QY 1261 GAGCTAACGTTTATGTATACAGAAATGCAGACATCGCTGCTTTACGTATCTAGTGG 1320
Db 1261 GAGCTAACGTTTATGTATACAGAAATGCAGACATCGCTGCTTTACGTATCTAGTGG 1320
QY 1321 GAGAGATGCAGAGACCGCTGTGTCTGCGGCTGTGAGCTGAGAGAGAGAGTGTATTTCC 1380
Db 1321 GAGAGATGCAGAGACCGCTGTGTCTGCGGCTGTGAGCTGAGAGAGAGTGTATTTCC 1380
QY 1381 TCCACACTGCGGAGAGAGATGAGACAACTATGATCATCATCTGTACGTACGTACG 1440
Db 1381 TCCACACTGCGGAGAGAGATGAGACAACTATGATCATCATCTGTACGTACGTACG 1440
QY 1441 ATACACAGAACTTCTATTAACAGTACAGTACGCTGCGGCTGCAATGGTGGCCAGTCACT 1500
Db 1441 ATACACAGAACTTCTATTAACAGTACAGTACGCTGCGGCTGCAATGGTGGCCAGTCACT 1500
QY 1501 TCTACACTGTCCAGTACAGACCTGCTGTTGACTGTGTATTCACGCTGTGTCATGCCA 1560
Db 1501 TCTACACTGTCCAGTACAGACCTGCTGTTGACTGTGTATTCACGCTGTGTCATGCCA 1560
QY 1561 TGTCTGAGACCCCACTTCAATGAAATCCCAACAGAGTCCGGTGAAGGGGAAATATCTCT 1620
Db 1561 TGTCTGAGACCCCACTTCAATGAAATCCCAACAGAGTCCGGTGAAGGGGAAATATCTCT 1620
QY 1621 TCACCTTCAGGGGAGAGAAATGAGTCTGTGAGTCTAGCTTGTAGAGGCGCGCTCCT 1680
Db 1621 TCACCTTCAGGGGAGAGAAATGAGTCTGTGAGTCTAGCTTGTAGAGGCGCGCTCCT 1680
QY 1681 TCGAAGAGAAATGAGAGGCGCACTCCCGGAGCTACGATGACCGGATCATTTGCCACCC 1740
Db 1681 TCGAAGAGAAATGAGAGGCGCACTCCCGGAGCTACGATGACCGGATCATTTGCCACCC 1740
QY 1741 TGAAGGCGGTGCAGAGACAGAAAGTGAATCAGTCCGTGTGGAATTCACCTGCAAAAACC 1800
Db 1741 TGAAGGCGGTGCAGAGACAGAAAGTGAATCAGTCCGTGTGGAATTCACCTGCAAAAACC 1800
QY 1801 AGCCCAAAACCAGCTGCGCACTGAGTGGCACTGTGTGAGAGCGGAGAGACGCTTGG 1860
Db 1801 AGCCCAAAACCAGCTGCGCACTGAGTGGCACTGTGTGAGAGCGGAGAGACGCTTGG 1860
QY 1861 AATGTGTAAGCTCTCCACTTTCGCTCATATTAACCCCGGGAGACCGTCTGTGTTA 1920
Db 1861 AATGTGTAAGCTCTCCACTTTCGCTCATATTAACCCCGGGAGACCGTCTGTGTTA 1920
QY 1921 TTTCTCTGAGGTGTGCAACAGGCTCTTGAAGCCCTGGAAGTGGGTGCGCTCCCGGTGG 1980
Db 1921 TTTCTCTGAGGTGTGCAACAGGCTCTTGAAGCCCTGGAAGTGGGTGCGCTCCCGGTGG 1980
QY 1981 TGTGTGGGAGAGAGTCCAGCTTCCCTACAGAGACATGCTGACATGAGAGAGGCGGCC 2040
Db 1981 TGTGTGGGAGAGAGTCCAGCTTCCCTACAGAGACATGCTGACATGAGAGAGGCGGCC 2040
QY 2041 TGTGTGTGCCAAAGCTGTGTGTTACCGAGGTTCATTTCTGTCTGAGAAGCTCTCCGATA 2100
|||||

Db 2041 TGTGTGTGCCAAAGCTGTGTGTTACCGAGGTTCATTTCTGTCTGAGAAGCTCTCCGATA 2100
QY 2101 GTGACCTCTGTGCTATGAGCGGCGCAAGGCGCTTTCTGTGGAGACTTACTTCCACTG 2160
Db 2101 GTGACCTCTGTGCTATGAGCGGCGCAAGGCGCTTTCTGTGGAGACTTACTTCCACTG 2160
QY 2161 CTGACAGTATTTTAAATACCCTGTGCTATGATTAAGAGCTGCATCCAGATCCAGCCG 2220
Db 2161 CTGACAGTATTTTAAATACCCTGTGCTATGATTAAGAGCTGCATCCAGATCCAGCCG 2220
QY 2221 CTCCCAATCGGGAGAGGGGCGAGTGAATCCCGCCCGCTTACGCAAGGCGGCTGGAA 2280
Db 2221 CTCCCAATCGGGAGAGGGGCGAGTGAATCCCGCCCGCTTACGCAAGGCGGCTGGAA 2280
QY 2281 CTGACCCCAACATGCTGACAAAGGGGAGCTGAGCTGGGGCCAGTGAAGAGAGCCGC 2340
Db 2281 CTGACCCCAACATGCTGACAAAGGGGAGCTGAGCTGGGGCCAGTGAAGAGAGCCGC 2340
QY 2341 CCTACGCTCCACCAATACCTCCGCAATTTCACTGTGACTGTCACTGACTTTTACCGCA 2400
Db 2341 CCTACGCTCCACCAATACCTCCGCAATTTCACTGTGACTGTCACTGACTTTTACCGCA 2400
QY 2401 GCTGNACTGTGCTCCAGGGGCTTTCCATCTTTTCCCGACACTCCCTTGAACCTGTGT 2460
Db 2401 GCTGNACTGTGCTCCAGGGGCTTTCCATCTTTTCCCGACACTCCCTTGAACCTGTGT 2460
QY 2461 TGCCTTCAGAGCCCAAAATCTTGGGCTCAGGAGTGGCTTTGGGCTATTTGGTGGTAG 2520
Db 2461 TGCCTTCAGAGCCCAAAATCTTGGGCTCAGGAGTGGCTTTGGGCTATTTGGTGGTAG 2520
QY 2521 CTGGGGGTTCTGGCAAGGAATTCAGGAGGCGCTTGGAGGCAATTTCCCGAGAGCACT 2580
Db 2521 CTGGGGGTTCTGGCAAGGAATTCAGGAGGCGCTTGGAGGCAATTTCCCGAGAGCACT 2580
QY 2581 TCACGGTGTGATGTTGACTTATGAGCGGAGAGAGTGGTATTAAGCTTTAGAGAGGC 2640
Db 2581 TCACGGTGTGATGTTGACTTATGAGCGGAGAGAGTGGTATTAAGCTTTAGAGAGGC 2640
QY 2641 TGAATGCGCTCCCTTACCTGAACAAAGTGTGGTGTGGAATTTCCCAAGCTGCCAT 2700
Db 2641 TGAATGCGCTCCCTTACCTGAACAAAGTGTGGTGTGGAATTTCCCAAGCTGCCAT 2700
QY 2701 CAGAGGACCTTCTGTGGGCTGACATGAGCTGGCTTCCATATGATGGTGTCCGTACTGAGAA 2760
Db 2701 CAGAGGACCTTCTGTGGGCTGACATGAGCTGGCTTCCATATGATGGTGTCCGTACTGAGAA 2760
QY 2761 ACAGTTTGAACAACGATCTTACCTGGAATGAAATGAGACAGAGGCACTCCGTCGA 2820
Db 2761 ACAGTTTGAACAACGATCTTACCTGGAATGAAATGAGACAGAGGCACTCCGTCGA 2820
QY 2821 TTGATGACGATGCTACCTCCGCAATGAGAAATCATGTTTGGTTCGGGTGTGAGAG 2880
Db 2821 TTGATGACGATGCTACCTCCGCAATGAGAAATCATGTTTGGTTCGGGTGTGAGAG 2880
QY 2881 AAGCTCGGGAGCGGATGCTGGGCTTCCCTGGCCCTTACAGCATGGGAGCATTCGCCATC 2940
Db 2881 AAGCTCGGGAGCGGATGCTGGGCTTCCCTGGCCCTTACAGCATGGGAGCATTCGCCATC 2940
QY 2941 AGTCTGGCTTACAACTCACTACTCTGTGAGTGTCCATGCTGTGACAGAGTGTG 3000
Db 2941 AGTCTGGCTTACAACTCACTACTCTGTGAGTGTCCATGCTGTGACAGAGTGTG 3000
QY 3001 CCTTTTTCACAACTATTAATGCTACCTGTATTTGATGATGAGTCCCGAGGCAATCCGGG 3060
Db 3001 CCTTTTTCACAACTATTAATGCTACCTGTATTTGATGATGAGTCCCGAGGCAATCCGGG 3060
QY 3061 ACATGAGTGAATGATTAATCACTGTGAGAGACATTTGCCATGAACTTCTGTCTCCACA 3120
Db 3061 ACATGAGTGAATGATTAATCACTGTGAGAGACATTTGCCATGAACTTCTGTCTCCACA 3120
QY 3121 TCACTGGAAGCCCGCCATCAAGGTGACTCAGGAGTGCATTCGATGCGCAGAGATGCC 3180
Db 3121 TCACTGGAAGCCCGCCATCAAGGTGACTCAGGAGTGCATTCGATGCGCAGAGATGCC 3180
```

(E-mail: cdna1nfo@kazusa.or.jp, Tel: +81-438-52-3913,
 Fax: +81-438-52-3914)
 2 (sites)
 Nagase, T., Ishikawa, K., Miyajima, N., Tanaka, A., Kotani, H.,
 Nomura, N. and Ohara, O.
 Prediction of the coding sequences of unidentified human genes. IX
 The complete sequences of 100 new cDNA clones from brain which can
 code for large proteins in this
 DNA Res. 5 (1), 31-39 (1998)
 98390545

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	FEATURES	SOURCE
0Y	1	GGCGGGTCCCTGAGCTGAGACCGCGAAGCAAGCCCTGGAGGTTCACTCTTCAAGAAAGT	60			
Db	23	GGCGGGTCCCTGAGCTGAGACCGCGAAGCAAGCCCTGGAGGTTCACTCTTCAAGAAAGT	82			
0Y	61	CGTGTGCTGAGAGTGTAATGCTACACAGTCAAGAGAAAGGAGGCTCTGAACACATGGC	120			
Db	83	CGTGTGCTGAGAGTGTAATGCTACACAGTCAAGAGAAAGGAGGCTCTGAACACATGGC	142			
0Y	121	CTGATTTGTTGGCAAGGCATCATAGAAGCTGGCATTTATTTGTTCTTAACCTATTACT	180			
Db	143	CTGATTTGTTGGCAAGGCATCATAGAAGCTGGCATTTATTTGTTCTTAACCTATTACT	202			
0Y	181	GTAATACGTGAATAGACATATGACATATTTGTTGTCAGCAAAACCAAGAAAGAGAGC	240			
Db	203	GTAATACGTGAATAGACATATGACATATTTGTTGTCAGCAAAACCAAGAAAGAGAGC	262			
0Y	241	TATGGCATTTGAAAAAGTCTGTGATATTCAGAGGTTTCTTCTGGGTTTCATCTACGAG	300			
Db	263	TATGGCATTTGAAAAAGTCTGTGATATTCAGAGGTTTCTTCTGGGTTTCATCTACGAG	322			
0Y	301	TACTCTCTCCCTTCACTTCAGCAAGAAATGTGGCAACTTTTATCGTTGGATTAAGATTAA	360			

BASE COUNT 1279 a 1687 c 1706 g 1517 t

ORIGIN

Query Match 99.9%; Score 6163.8; DB 9; Length 6189;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 6165; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1. 6189 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="HG1237"
 /sex="male"
 /tissue_type="brain"
 /clone_id="pbluescriptII SK plus"
 /gene="K1AA0519"
 616..3376
 /gene="K1AA0519"
 616..3375
 /codon_start=1
 /product="K1AA0519 protein"
 /protein_id="BA25445.1"
 /db_xref="GI:3043562"
 /translation="MTGYTLMRNGAGNGGNGGOMLRWSNRIRLWLSFTLFLVILVPEP
 LLAHYLLTLDEADBAKGRIFRPGYNGLGKLVKYLDCRIRESVSEELQLKRRGE
 LNSLEAIKLTKEACKSLENAKODLQKLVISOTESHSTVELQAQNPRLSPRL
 PKDAGLDPKATRGCLNLCFDSIRCPVLSGFPVYVYSDDFGFSYLDPLVKQF
 OKPAARAVYENADICLVYLVGEMQEPVLPABELEKQLVLPMPRTDGNHVIY
 NLSKSDONLNVSTSRAMVAOSTFTVYVRGDFLVSPVLAHASEBNPEITPO
 VVKKRYLFTTEGKEIESLRSLQEARFEEEMGDDPADVDORITATIKAAODSKLD
 OVLFTCKNQPKPSLPTFEMALGGERBRELKLTSMPLALITPGDRLVYSGCATR
 LEAEVAEAVPVLEQVQLVQPLMLQMWNAIVPPRTVEVHFILRSLSDLLAM
 ROGELEWETVSTADSIENFTVLMIRRIQIPAPLTREAEAEIPHRSGAAGDPN
 MADNDLIDGVEETPEPTASPRI LNFELVTVDEYRSMNCAPGPFILFPTPDVPL
 SEAKTIGCTGFRPLIGGAGSGSEKFEQALAGVNPBQFVVMILYEREELVNSLER
 LNGDPLYNKVVVYVNSPKPLSPEDLMDIGVPIVYVTEKNSLNRRFLPNEETEFAI
 LSIDDAHLRHDEIMFGFRVWRREARDIVGPGRYHAMDIPQSMVLYNSYSELSM
 LTGAIFKRYAYLVSYMPQAIARDNDYEDINCEDIAMNPLVSHITKRP IAKTYSMT
 FRCPCQPALSHDSDSHFERHKCINFPVKYKGMPLLYTFQFVDSVLEKTRLPDHTK
 CFKPL"

Db 323 TACCTCTCCCTTTTCAATCTCAGCAAGATGTGGCACCTTTTATCGTTTGATTAAGATTAA 382
OY 361 GCACATGTTCTTGTGTCAACAGCCAGAACTTAATAATCTGCTGGAAATAGGTCAGAGACCA 420
Db 383 GGCATGTTCTTGTGTCAACAGCCAGAACTTAATAATCTGCTGGAAATAGGTCAGAGACCA 442
OY 421 TTTTCACTGCACTGAGAGAAATGAATGTTCAATTTTATTTGGTGGCTTGTCTGGGAGC 480
Db 443 TTTTCACTGCACTGAGAGAAATGAATGTTCAATTTTATTTGGTGGCTTGTCTGGGAGC 502
OY 481 ACCTTAACCTTCTGGAAACGTGTCAAGTAAGAGATCGTTTGTGGAAATGCAACCC 540
Db 503 ACCTTAACCTTCTGGAAACGTGTCAAGTAAGAGATCGTTTGTGGAAATGCAACCC 562
OY 541 ATGATTATGGAGATGACCCAGAGTATCTGGGGGAGAGCTGACAGAGACTATGACAG 600
Db 563 ATGGTTATGGCGAGTGAACCCAGAGTATCTGGGGGAGAGCTGACAGAGACTATGACAG 622
OY 601 GCTATACATGCTGCGGAATGGGGGCGGGGGAAGCGAGGTCAGACCTGATCTCGCGT 660
Db 623 GCTATACATGCTGCGGAATGGGGGCGGGGGAAGCGAGGTCAGACCTGATCTCGCGT 682
OY 661 GGTCAACCCGATCGCGCTACGTTGGCTCAGCTTCAAGCTTTTGTATCTGCTGTCT 720
Db 683 GGTCAACCCGATCGCGCTACGTTGGCTCAGCTTCAAGCTTTTGTATCTGCTGTCT 742
OY 721 TCCCGCTCATCGCCACACTATTACCTCAGCACTCTGGATGAGGCTGATGAGGCAAGCA 780
Db 743 TCCCGCTCATCGCCACACTATTACCTCAGCACTCTGGATGAGGCTGATGAGGCAAGCA 802
OY 781 GGAATTTTGGTCCCCGGGTGGGGAAGAGCTGTGCGAGGTGAAGACAGCTGTCTGATCTGT 840
Db 803 GGAATTTTGGTCCCCGGGTGGGGAAGAGCTGTGCGAGGTGAAGACAGCTGTCTGATCTGT 862
OY 841 GCGCATCCGCGAGTGGGTAGTGAAGAGCTCTTGCAGCTGAGAGCCAGCAAGCCAGAGC 900
Db 863 GCGCATCCGCGAGTGGGTAGTGAAGAGCTCTTGCAGCTGAGAGCCAGCAAGCCAGAGC 922
OY 901 TGAACAGCAGATCGCCAACTGAATCTGAAGATCGAAGCCCTGTAAGAGAGCAATTGAGA 960
Db 923 TGAACAGCAGATCGCCAACTGAATCTGAAGATCGAAGCCCTGTAAGAGAGCAATTGAGA 982
OY 961 ACCCGAAGCAGAGACTGTCTCAAGTAATGTCAATCAGCCAGAGCCAGACTTCTTACA 1020
Db 983 ACCCGAAGCAGAGACTGTCTCAAGTAATGTCAATCAGCCAGAGCCAGACTTCTTACA 1042
OY 1021 AGGAGCTGATGGCCAGAAACAGCCAGAGCTGCTGCCATCCGAGCTCTCCAGAGA 1080
Db 1043 AGGAGCTGATGGCCAGAAACAGCCAGAGCTGCTGCCATCCGAGCTCTCCAGAGA 1102
OY 1081 AGGAGCTGATGGCCAGAAACAGCCAGAGCTGCTGCCATCCGAGCTCTCCAGAGCTGT 1140
Db 1103 AGGAGCTGATGGCCAGAAACAGCCAGAGCTGCTGCCATCCGAGCTCTCCAGAGCTGT 1162
OY 1141 TTTGATTTATCTGTGCTTGTGCTCTCACTCTGCTGCTTCCGGCTTCAAGCTATGACAGT 1200
Db 1163 TTTGATTTATCTGTGCTTGTGCTCTCACTCTGCTGCTTCCGGCTTCAAGCTATGACAGT 1222
OY 1201 AGTTTGCTTTGGCAGTACCTGATCCCTTGGTCAAGAGGCTTTTTCAGGCGACAGCAC 1260
Db 1223 AGTTTGCTTTGGCAGTACCTGATCCCTTGGTCAAGAGGCTTTTTCAGGCGACAGCAC 1282
OY 1261 GAGCTAACGTTTATGTTACAGAAATGACAGATCGCTTCTTACAGTATCTAGTGG 1320
Db 1283 GAGCTAACGTTTATGTTACAGAAATGACAGATCGCTTCTTACAGTATCTAGTGG 1342
OY 1321 GAGAGATGACAGAGCCGCTGCTGCGCTCTGAGCTGAGAGAGCAATTTATTTCC 1380
Db 1343 GAGAGATGACAGAGCCGCTGCTGCGCTCTGAGCTGAGAGAGCAATTTATTTCC 1402
OY 1381 TGGCACAATGCGGAGAGAGTGAACACCATGTCATCATCAATTCGTACATGTAAGT 1440
Db 1403 TGGCACAATGCGGAGAGAGTGAACACCATGTCATCATCAATTCGTACATGTAAGT 1462

OY 1441 ATACACAGAACCTTCTATTAACGTCAAGTATGCGCCGTGCCATGTTGGCCAGTCCACT 1500
Db 1463 ATACACAGAACCTTCTATTAACGTCAAGTATGCGCCGTGCCATGTTGGCCAGTCCACT 1522
OY 1501 TCTACACTGTCCAGTATACAACTGGCTTTGACTTGGTCTATTCACCGCTGTGTCAATGCCA 1560
Db 1523 TCTACACTGTCCAGTATACAACTGGCTTTGACTTGGTCTATTCACCGCTGTGTCAATGCCA 1582
OY 1561 TGTGTGAGCCCAACTCATGGAATCCACACAGAGTGGCGGGGAAGGGAATATCTCT 1620
Db 1583 TGTGTGAGCCCAACTCATGGAATCCACACAGAGTGGCGGGGAAGGGAATATCTCT 1642
OY 1621 TCACCTTCCAGGGCGAGAAAGATTGAGTCTGTAGGTCTAGCTTTCAGAGGCCGCTCT 1680
Db 1643 TCACCTTCCAGGGCGAGAAAGATTGAGTCTGTAGGTCTAGCTTTCAGAGGCCGCTCT 1702
OY 1681 TCGAAGAGAAATGAGAGGCGAGACCTCCCGCCGACATACGATACCGAGATTCGCACCC 1740
Db 1703 TCGAAGAGAAATGAGAGGCGAGACCTCCCGCCGACATACGATACCGAGATTCGCACCC 1762
OY 1741 TGAAGCGGTGCGAGACAGCAAGCTGATCAGTCTGTGTGAAATTCACCTGCAAAAACC 1800
Db 1763 TGAAGCGGTGCGAGACAGCAAGCTGATCAGTCTGTGTGAAATTCACCTGCAAAAACC 1822
OY 1801 AGCCCAACCCAGCCTGCCAGCTGAGTGGGCACTGTGTGAGAGCGGGAGAGACCGCTTGG 1860
Db 1823 AGCCCAACCCAGCCTGCCAGCTGAGTGGGCACTGTGTGAGAGCGGGAGAGACCGCTTGG 1882
OY 1861 AATTGCTGAAGCTCTCCACTTCCGCTCATCATTTACCCCGGGGAGCCCTGCTTGGTTA 1920
Db 1883 AATTGCTGAAGCTCTCCACTTCCGCTCATCATTTACCCCGGGGAGCCCTGCTTGGTTA 1942
OY 1921 TTTCTCTGTGGTGTGCAACACAGGCTTTCGAAGCCCTGGAAGTGGTGGCTCCGCTGG 1980
Db 1943 TTTCTCTGTGGTGTGCAACACAGGCTTTCGAAGCCCTGGAAGTGGTGGCTCCGCTGG 2002
OY 1981 TGTGGGGAGCAGGTTCAGCTTCCCTACCAAGACATGCTGCAAGTGGAAAGAGGCGGCC 2040
Db 2003 TGTGGGGAGCAGGTTCAGCTTCCCTACCAAGACATGCTGCAAGTGGAAAGAGGCGGCC 2062
OY 2041 TGTGGTGGCAAGGCTCGGTACCGAGGTTTCTCGCTCAGAAAGCTCTCCGATA 2100
Db 2063 TGTGGTGGCAAGGCTCGGTACCGAGGTTTCTCGCTCAGAAAGCTCTCCGATA 2122
OY 2101 GTGACCTCTGTGCTATGAGGCGGCAAGGCGCTTCTCTGGAGACTTACTTCTCCACTG 2160
Db 2123 GTGACCTCTGTGCTATGAGGCGGCAAGGCGCTTCTCTGGAGACTTACTTCTCCACTG 2182
OY 2161 CTGACAGTATTTTAAATACCGTGTGCTATGATGATGAGACTGCGATCCAGATCCAGCCG 2220
Db 2183 CTGACAGTATTTTAAATACCGTGTGCTATGATGATGAGACTGCGATCCAGATCCAGCCG 2242
OY 2221 CTGCCATCCGGGGAAGAGGGGCGAGCTGAGATCCCCACCGTTTACAGCAAGCGGCTGGAA 2280
Db 2243 CTGCCATCCGGGGAAGAGGGGCGAGCTGAGATCCCCACCGTTTACAGCAAGCGGCTGGAA 2302
OY 2281 CTGACCCCAACATGCTGACAAAGGGGAGCTGAGACTTGGGGCCAGTGGAGAGCAGACCCG 2340
Db 2303 CTGACCCCAACATGCTGACAAAGGGGAGCTGAGACTTGGGGCCAGTGGAGAGCAGACCCG 2362
OY 2341 CTTAGCCTTCACCCAGATACCTCCGCAATTTCACTGTGACTGTCACTGACTTTTACCGCA 2400
Db 2363 CTTAGCCTTCACCCAGATACCTCCGCAATTTCACTGTGACTGTCACTGACTTTTACCGCA 2422
OY 2401 GCTGGAAGTGTGTCCAGGGCTTTTCCATCTTTTCCCCCACTCTCCCTTTGACACCTGTGT 2460
Db 2423 GCTGGAAGTGTGTCCAGGGCTTTTCCATCTTTTCCCCCACTCTCCCTTTGACACCTGTGT 2482
OY 2461 TGGCCTCAGAGGCAAAATTTTGGGCTCAGGAGCTGGCTTTTGGGCTATTTGGGTGGAG 2520
Db 2483 TGGCCTCAGAGGCAAAATTTTGGGCTCAGGAGCTGGCTTTTGGGCTATTTGGGTGGAG 2542

QY 2521 CTGGGGGTTCTGCGAAGAAATTTCCAGCAGCGCTTGGAGCAATGTTCCTCCGAGAGCACT 2580
|||||
Db 2543 CTGGGGGTTCTGCGAAGAAATTTCCAGCAGCGCTTGGAGCAAGTTCCCGAGAGCACT 2602
QY 2581 TCACGGTGTGATGTGACTTATGAGCGGAGAGAGTCTTATGAATCTTTTGGAGGCG 2640
|||||
Db 2603 TCACGGTGTGATGTGACTTATGAGCGGAGAGAGTCTTATGAATCTTTTGGAGGCG 2662
QY 2641 TGAATGGCCCTCCCTTACCTCGAACAAGTCTGTGTGGAAATTTCTCCAGCTGCCAT 2700
2663 TGAATGGCCCTCCCTTACCTCGAACAAGTCTGTGTGGAAATTTCTCCAGCTGCCAT 2722
QY 2701 CAGAGGACCTTCTGTGGCTGACATTTGGCGTTCCATCATGAGTGGTCCGCTACGAGAA 2760
|||||
Db 2723 CAGAGGACCTTCTGTGGCTGACATTTGGCGTTCCATCATGAGTGGTCCGCTACGAGAA 2782
QY 2761 ACAGTTTGAACAACCGATTTCTTACCCTGGAATGAAATGAGACAGAGGCAATCTGTCCA 2820
2783 ACAGTTTGAACAACCGATTTCTTACCCTGGAATGAAATGAGACAGAGGCAATCTGTCCA 2842
QY 2821 TTGATGACGATCTGACCTCGCGCAGAGAAATCATGTTTGGGTTCCGGGTGTGGAGAG 2880
2843 TTGATGACGATCTGACCTCGCGCAGAGAAATCATGTTTGGGTTCCGGGTGTGGAGAG 2902
QY 2881 AAGCTCGGAGCCGATCTGTGGCTTCCCTGGCCGTTTACACGCAATGGAGCATCCCCCATC 2940
2903 AAGCTCGGAGCCGATCTGTGGCTTCCCTGGCCGTTTACACGCAATGGAGCATCCCCCATC 2962
QY 2941 AGCTCGGCTCTACAACTCCAACTACTCTGTGAGCTGTCCATGGTGTGACAGGTGTG 3000
2963 AGCTCGGCTCTACAACTCCAACTACTCTGTGAGCTGTCCATGGTGTGACAGGTGTG 3022
QY 3001 CCTTCTTACAAATATATGCTACCTACCTGATTTCTTATGATGATGCCAGCCATCCGGG 3060
3023 CCTTCTTACAAATATATGCTACCTACCTGATTTCTTATGATGATGCCAGCCATCCGGG 3082
QY 3061 ACATGTGTGATGATATACATCAACTGTGAGACATTTGCCATGAATCTTCTTGTCCACA 3120
3083 ACATGTGTGATGATATACATCAACTGTGAGACATTTGCCATGAATCTTCTTGTCCACA 3142
QY 3121 TCACCTCGAAGCCCGCATTAAGTGTACCTGACGCTGAGCATTCGATGGCCAGATGCC 3180
3143 TCACCTCGAAGCCCGCATTAAGTGTACCTGACGCTGAGCATTCGATGGCCAGATGCC 3202
QY 3181 CTCAGGCCCTGTCTCATGATGATCCCACTTCCACGAGGGGCAACATGATCAACTTCT 3240
3203 CTCAGGCCCTGTCTCATGATGATCCCACTTCCACGAGGGGCAACATGATCAACTTCT 3262
QY 3241 TCGTGAAGGTGTAGGCTACATGCCCCCTCTGTATACAGCAAGTTCAGAGGTGATTTGTGC 3300
3263 TCGTGAAGGTGTAGGCTACATGCCCCCTCTGTATACAGCAAGTTCAGAGGTGATTTGTGC 3322
QY 3301 TCTTCAAGACAGCCTGCCCCATGACAAAGCAAGTGTTCATGTTCAATCTATGAGGCGACG 3360
3323 TCTTCAAGACAGCCTGCCCCATGACAAAGCAAGTGTTCATGTTCAATCTATGAGGCGACG 3382
QY 3361 GCAAGGCTGTGGGAAAGAGATGAGCAGAGGAGAGAGTGGTCCCAAGGTTCTTAGGCA 3420
3383 GCAAGGCTGTGGGAAAGAGATGAGCAGAGGAGAGAGTGGTCCCAAGGTTCTTAGGCA 3442
QY 3421 TTGCAAGACCTTGGGACATCTGTGTGTGTGGGCCAGAGCCTGTCTGGAAGGGGAGC 3480
3443 TTGCAAGACCTTGGGACATCTGTGTGTGTGGGCCAGAGCCTGTCTGGAAGGGGAGC 3502
QY 3481 CAGGAGGATGGAAGAAACCGCTCTCTTATCTTGAAGTACGCCACATGAGGCTGTGAG 3540
3503 CAGGAGGATGGAAGAAACCGCTCTCTTATCTTGAAGTACGCCACATGAGGCTGTGAG 3562
QY 3541 CCTGGGGGAGATCCCGGGGTTCCCAACAGGCGCACTGACTGATTAACACTGAGG 3600
3563 CCTGGGGGAGATCCCGGGGTTCCCAACAGGCGCACTGACTGATTAACACTGAGG 3622
QY 3601 ACTGTGGGACTGTGCAAGTCACTACACCGTTCTGTACGCCAGACAGTGTGGTCTGTG 3660
|||||

Db 3623 ACTGTGGGACTGTGCAAGTCACTACACCGTTCTGTACGCCAGACAGCACTGTGTGTG 3682
QY 3661 GTTTTACATTCATAAACAATATATGATTTATTTAAAAAGAGAGTTTCAATTTGGC 3720
3683 GTTTTACATTCATAAACAATATATGATTTATTTAAAAAGAGAGTTTCAATTTGGC 3742
QY 3721 ATTCAAGGCTTATTTATATATATGTGTGTATATATATATATATATATATATATAT 3780
3743 ATTCAAGGCTTATTTATATATATGTGTGTATATATATATATATATATATATATATAT 3802
QY 3781 TATATATTTTGGCTGGGAGAGTGTGTTGGCTTTCTAAGGAGGAGACCGCGAGC 3840
3803 TATATATTTTGGCTGGGAGAGTGTGTTGGCTTTCTAAGGAGGAGACCGCGAGC 3862
QY 3841 TCCCTTGTCTGTATTTCTGCGAGAGATGGGCTGTGGCTTGTGTACATGCTTATCTTA 3900
3863 TCCCTTGTCTGTATTTCTGCGAGAGATGGGCTGTGGCTTGTGTACATGCTTATCTTA 3922
QY 3901 AAGATCATCTCCCATCTCCAGGCCATCTGTGTGACAGCAACAGAAAGGATGAAT 3960
3923 AAGATCATCTCCCATCTCCAGGCCATCTGTGTGACAGCAACAGAAAGGATGAAT 3982
QY 3961 TGGCCCTTGTGGGGGCTGAGCAAGGTCCTTCTTACCTTCTGTGCTTGTGCAATGCA 4020
3983 TGGCCCTTGTGGGGGCTGAGCAAGGTCCTTCTTACCTTCTGTGCTTGTGCAATGCA 4042
QY 4021 CCTGTATCATATTTCTTCTCCAGTGAATCCCTGAGGAGGCTGTGACCTGTGGGGCT 4080
4043 CCTGTATCATATTTCTTCTCCAGTGAATCCCTGAGGAGGCTGTGACCTGTGGGGCT 4102
QY 4081 TCAGCTTCTGCTGCTGGGGGCAAGCATTTTGTAGAGATTTATCTTATAGCCAGGCTTGC 4140
4103 TCAGCTTCTGCTGCTGGGGGCAAGCATTTTGTAGAGATTTATCTTATAGCCAGGCTTGC 4162
QY 4141 TCGTACTTATCCCTGCTCCCAATTTCTCTTGTGTGAGAGATTTAGAGAGCAAG 4200
4163 TCGTACTTATCCCTGCTCCCAATTTCTCTTGTGTGAGAGATTTAGAGAGCAAG 4222
QY 4201 AGTGAAGAAATATAGGGGCTGGAAGCGCACTCCAGATGGCTTCTTATCCCTGCTT 4260
4223 AGTGAAGAAATATAGGGGCTGGAAGCGCACTCCAGATGGCTTCTTATCCCTGCTT 4282
QY 4261 CTGTGTAACACAGTGTGTGGGCTCAGGCGCTTCTGAAAGTCTTCTTGTGATGG 4320
4283 CTGTGTAACACAGTGTGTGGGCTCAGGCGCTTCTGAAAGTCTTCTTGTGATGG 4342
QY 4321 ACAGGAGATCAGCAGGTGCATCTGTGTGTGAAAGTGTGTGCAAGTCAAGCTCC 4380
4343 ACAGGAGATCAGCAGGTGCATCTGTGTGTGAAAGTGTGTGCAAGTCAAGCTCC 4402
QY 4381 TCTCCCTAGTGAAGACAGCAAGTGTCTTGTGAGAAACCAACCGGCTGGCCGGGAAT 4440
4403 TCTCCCTAGTGAAGACAGCAAGTGTCTTGTGAGAAACCAACCGGCTGGCCGGGAAT 4462
QY 4441 TTTACAGCAAGCGCTGCTTGGATTAATCTTGGTGAATTTACCTTCCCCCGCT 4500
4463 TTTACAGCAAGCGCTGCTTGGATTAATCTTGGTGAATTTACCTTCCCCCGCT 4522
QY 4501 CTGTGTGAGCCCATCTGTGTATCTGTGTGTTTGGACCCCTAATGTCACTGTGCT 4560
4523 CTGTGTGAGCCCATCTGTGTATCTGTGTGTTTGGACCCCTAATGTCACTGTGCT 4582
QY 4561 GTAGAGTCCCCGAGGTTTGGATGTGTGTAGAACATGAGGAGGTGATTTGTGTGTA 4620
4583 GTAGAGTCCCCGAGGTTTGGATGTGTGTGTAGAACATGAGGAGGTGATTTGTGTGTA 4642
QY 4621 AGCTCATATCCAGCCTTGGATTTAAGGAGATTCACACCCGAGTTACACTTCCACT 4680
4643 AGCTCATATCCAGCCTTGGATTTAAGGAGATTCACACCCGAGTTACACTTCCACT 4702
QY 4681 CCTGTGTAGGATTTCTGTCCCTGGGCTGAAGCAATGAATATTTTGGGTGACG 4740
|||||

Db	4703	CCGTCCTAGAGTATTCGTGTTCCCTGGGGCTGAACGTAAATAGCTAAATTTTGGGTACG	4762
Oy	4741	GTGCGACAGTGGGAACTAGAGGGGTGTAGTGGCATTTGTTCAGGGATTTAGCCCATGAC	4800
Db	4763	GTGCGACATAGGGGAACTTAGAGGGGTGTAGTGGCATTTGTTCAGGGATTTAGCCCATGAC	4822
Oy	4801	GTGTTTCTTAACCCATCTTCTTGGAAGTGGAGTTGACCTGTGGAGTTTCTTACCAACTG	4860
Db	4823	GTGTTTCTTAACCCATCTTCTTGGAAGTGGAGTTGACCTGTGGAGTTTCTTACCAACTG	4882
Oy	4861	AACAAAAGCTCAGATTGTCCTGGTATGCACATGCGCTTAAGCCAGTTCCGTCCTCCCTA	4920
Db	4883	AACAAAAGCTCAGATTGTCCTGGTATGCACATGCGCTTAAGCCAGTTCCGTCCTCCCTA	4942
Oy	4921	GACCTTGGACCTCTGTGCTCTTATTTCTTGAAATACGTTCTCTCTGACCTGCTGTACC	4980
Db	4943	GACCTTGGACCTCTGTGCTCTTATTTCTTGAAATACGTTCTCTCTGACCTGCTGTACC	5002
Oy	4981	ACGTGGGGCCCTCTCAAGTACGTGTTTTGAAAGTCGGGCTCTTTTGTTAGTCCCAACCCAC	5040
Db	5003	ACGTGGGGCCCTCTCAAGTACGTGTTTTGAAAGTCGGGCTCTTTTGTTAGTCCCAACCCAC	5062
Oy	5041	CTGTAGGGCTAGCTCGGGCTTAAAGGAACTCTCCCATTTGGCAACCGAACCCGGGCGCG	5100
Db	5063	CTGTAGGGCTAGCTCGGGCTTAAAGGAACTCTCCCATTTGGCAACCGAACCCGGGCGCG	5122
Oy	5101	CCAGAGCTGTGTTTCCAAAGTTCCCGGCGCCCAACCCAGCATGAGCCTGTAGTCCCC	5160
Db	5123	CCAGAGCTGTGTTTCCAAAGTTCCCGGCGCCCAACCCAGCATGAGCCTGTAGTCCCC	5182
Oy	5161	TGCTGAGGCACTGTGGTTATGTTCCCAAGCATGTGGGGTGTAGACGCCCTTCTCAGAACTT	5220
Db	5183	TGCTGAGGCACTGTGGTTATGTTCCCAAGCATGTGGGGTGTAGACGCCCTTCTCAGAACTT	5242
Oy	5221	TCTAGTTGCCCTCTACCTGACCTCTGACCTGTATTCCTTTTGTAGAGTACCTCTTCCTCT	5280
Db	5243	TCTAGTTGCCCTCTACCTGACCTCTGACCTGTATTCCTTTTGTAGAGTACCTCTTCCTCT	5302
Oy	5281	CGGGGAGCCAAAGAGTGTGTGTGTGGCGCTATATTGTGGCTGCTCAATTTCACTGTGTTTC	5340
Db	5303	CGGGGAGCCAAAGAGTGTGTGTGTGGCGCTATATTGTGGCTGCTCAATTTCACTGTGTTTC	5362
Oy	5341	TTTTTAATGTAGGAATCTACATCTGTGACTTCACTGGGACTCGGTGTAGCCGGGCGCTGTG	5400
Db	5363	TTTTTAATGTAGGAATCTACATCTGTGACTTCACTGGGACTCGGTGTAGCCGGGCGCTGTG	5422
Oy	5401	TGTGGTGGAGCCCCCTTACGGGGACACGTAGCTGTGGGGCGCTGTGTGTGTGTGGAGCA	5460
Db	5423	TGTGGTGGAGCCCCCTTACGGGGACACGTAGCTGTGGGGCGCTGTGTGTGTGTGGAGCA	5482
Oy	5461	GGGGCTCTCCCTTTAGTGTAGGAGCCAGGTGTGTGGGGCCCCCAATGTCACTGTGTGATCTAAG	5520
Db	5483	GGGGCTCTCCCTTTAGTGTAGGAGCCAGGTGTGTGGGGCCCCCAATGTCACTGTGTGATCTAAG	5542
Oy	5521	AAGGGCTGAGTGTCTGTACACCAAAACATGTCCGCAAGGAGGCTGTGTGTCGGGTCTTC	5580
Db	5543	AAGGGCTGAGTGTCTGTACACCAAAACATGTCCGCAAGGAGGCTGTGTGTCGGGTCTTC	5602
Oy	5581	CAACAAGACAGCCCTCTTGACCCCTTGAAAGAACTGTGGCTTGAAGACATGTGCACACAGG	5640
Db	5603	CAACAAGACAGCCCTCTTGACCCCTTGAAAGAACTGTGGCTTGAAGACATGTGCACACAGG	5662
Oy	5641	CTTGTAGGGGACACCCCTCTCTGACGAGAGAGCAGCAAGTGTGGCCACAGTGTACATGTGCA	5700
Db	5663	CTTGTAGGGGACACCCCTCTCTGACGAGAGAGCAGCAAGTGTGGCCACAGTGTACATGTGCA	5722
Oy	5701	GGTGTCTTCTCACACGGGAAAGCCGCGACCTGTGTGACTGCTTGAATGTGGAAAGCGCG	5760
Db	5723	GGTGTCTTCTCACACGGGAAAGCCGCGACCTGTGTGACTGCTTGAATGTGGAAAGCGCG	5782
Oy	5761	CCACAGACCCCGGGTCTCTCTTGGCTGTCTGTGGGGCGCCCTGTGGCCACTTGTCTGGCT	5820
Db	5783	CCACAGACCCCGGGTCTCTCTTGGCTGTCTGTGGGGCGCCCTGTGGCCACTTGTCTGGCT	5842

[illegible]

MADNGDLIDLPVEEPYASRYLRNPLTVDPYISMNCAPEPHLEPHTPDVPLP
SEAKPLIGSGTGRPIGIGGAGSGSKEROALIGVNPBQFTVMITYEREVLNLSLER
LNLPLINKYVYVYVNSPKLPSEDLPLMDVIGPIMVYFTEKNSJNNRFLPWNIELEAI
LSTDDAHHLHDEIMFPGFVWRARORIVGFPARIVAMDIPHSMLYNSMYSELESMY
LTGAAFHHKYYALYSYVMPQALRDYVDEYINCEIDJAMNLFVSHITRKPRIKYSMT
FRPCGPQALSHSDSHFHERHKCINFEVYVGMPLLYIQFRNDSVLEFTRLPDHDKT
CPKFI¹

BASE COUNT 1261 a 1681 c 1694 g 1511 t
ORIGIN

Query Match 99.2%; Score 6124.8; DB 9; Length 6147;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6137; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GGGGGTCCCTGAGCTGGAGGCGGAGAGACCCCTGAGGTTCACTCTTCAAGAAAGT 60
DB 9 GGGGGTCCCTGAGCTGGAGGCGGAGAGACCCCTGAGGTTCACTCTTCAAGAAAGT 68
QY 61 CGTGTGCTAGAGTGTATCTACACAAGTCAGAGGAAGGAGGCTCTGAACACATGGC 120
DB 69 CGTGTGCTAGAGTGTATCTACACAAGTCAGAGGAAGGAGGCTCTGAACACATGGC 128
QY 121 CTGATTGTTGGCAAGGCAATCAAGAACTGGCAATTTATTTCTGTTCACTATTACT 180
DB 129 CTGATTGTTGGCAAGGCAATCAAGAACTGGCAATTTATTTCTGTTCACTATTACT 188
QY 181 GTATTAAGTGTGAATGACACTATGATATTTGTTGTCAGCAAAACCAAGAAAGAGC 240
DB 189 GTATTAAGTGTGAATGACACTATGATATTTGTTGTCAGCAAAACCAAGAAAGAGC 248
QY 241 TATGSCATTTGAAAAAGTCTGTCTGATTCAGAGGTTTTCCTGGGTTCTCATCAGG 300
DB 249 TATGSCATTTGAAAAAGTCTGTCTGATTCAGAGGTTTTCCTGGGTTCTCATCAGG 308
QY 301 TACCTCCCTCCCTTCATCTCAAGCAAGAAATGTGGACCTTTATCTGTTGATTAAGATTAA 360
DB 309 TACCTCCCTCCCTTCATCTCAAGCAAGAAATGTGGACCTTTATCTGTTGATTAAGATTAA 368
QY 361 GGACATGTTCTTGGTCAACAGCCAGAACTTAAATCTGCGGAATAGGGTCAGAGACCA 420
DB 369 GGACATGTTCTTGGTCAACAGCCAGAACTTAAATCTGCGGAATAGGGTCAGAGACCA 428
QY 421 TTTTACGTCGACGTGAGGAAAAATGAATGTTCATTTTATTGGAGCCCTGTCTGGAGAC 480
DB 429 TTTTACGTCGACGTGAGGAAAAATGAATGTTCATTTTATTGGAGCCCTGTCTGGAGAC 488
QY 481 ACACCTACTCTTCTGGAAGAGTGTCACTGAACAAGAGATCGTTTGTGGAATAGCAACC 540
DB 489 ACACCTACTCTTCTGGAAGAGTGTCACTGAACAAGAGATCGTTTGTGGAATAGCAACC 548
QY 541 ATGGTTATGCGAGATGCCGAGATGATCTGGGGGGGAGGCTGACAGAGACTCATYACAG 600
DB 549 ATGGTTATGCGAGATGCCGAGATGATCTGGGGGGGAGGCTGACAGAGACTCATYACAG 608
QY 601 GCTATACCATGCTGCGGAATGGGGGGGCGGGAAGGAGGTCAAGACTGATCTGGCT 660
DB 609 GCTATACCATGCTGCGGAATGGGGGGGCGGGAAGGAGGTCAAGACTGATCTGGCT 668
QY 661 GGTCACACCGCATCCGCTTACGTCGCTCAAGCTTCTTGTATCTCTGCTTCT 720
DB 669 GGTCACACCGCATCCGCTTACGTCGCTCAAGCTTCTTGTATCTCTGCTTCT 728
QY 721 TCCCGCTCATGCGCCCACTTACCTACACACTTGTGATAGGCTGATGAGGAGGCAAGC 780
DB 729 TCCCGCTCATGCGCCCACTTACCTACACACTTGTGATAGGCTGATGAGGAGGCAAGC 788
QY 781 GGATTTTGGTCCCGGGGTGGGAACGAGCTGTGCGAGGTGAAGCAAGCTGCTGATCT 840
DB 789 GGATTTTGGTCCCGGGGTGGGAACGAGCTGTGCGAGGTGAAGCAAGCTGCTGATCT 848
QY 841 GCCGATCCGGAGTGGGTGAGTGAAGACTCTGACAGCTGAGGCCAAGCCCAAGAGC 900

DB 849 GCCGATCCGGAGTGGGTGAGTGAAGAGCTCTTCAAGCTGAGAGCCCAAGCCCAAGAGC 908
QY 901 TGAACACGAGATGCGCAAGCTGATCTGAAGATCGAAGCCCTGATAGAAGCATTTAGA 960
DB 909 TGAACACGAGATGCGCAAGCTGATCTGAAGATCGAAGCCCTGATAGAAGCATTTAGA 968
QY 961 ACCGACAGAGAGCTCTGACCTGCAAGAAATGTGATCAGCAGAGCCGAGCATTTCTACA 1020
DB 969 ACCGACAGAGAGCTCTGACCTGCAAGAAATGTGATCAGCAGAGCCGAGCATTTCTACA 1028
QY 1021 AGAAGCTCATGCGCCAGAAACACAGCCCAAGCTGTCCGCTCATCCGACAGCTCCGAGAG 1080
DB 1029 AGAAGCTCATGCGCCAGAAACACAGCCCAAGCTGTCCGCTCATCCGACAGAG 1088
QY 1081 AGAAGCATGCGCGCTCTCCCTCCCGAAGGCCACTGCGGGCTGCGCGCTACACACTGCT 1140
DB 1089 AGAAGCATGCGCGCTCTCCCTCCCGAAGGCCACTGCGGGCTGCGCGCTACACACTGCT 1148
QY 1141 TTGATTAATCTGCTTCCCTCTCACTCTGAGGCTTCCGGTCTTACGCTTATGACAGTGACC 1200
DB 1149 TTGATTAATCTGCTTCCCTCTCACTCTGAGGCTTCCGGTCTTACGCTTATGACAGTGACC 1208
QY 1201 AGTTGTCTTGGCAGACTACCTGATCCCTTGTGCAAGCAGGCTTTTACGGCAGACGAC 1260
DB 1209 AGTTGTCTTGGCAGACTACCTGATCCCTTGTGCAAGCAGGCTTTTACGGCAGACGAC 1268
QY 1261 GAGCTAAGCTTATATGTTACAGAAAAATGACAGATGCGCTTCTTACGTATGATAGTG 1320
DB 1269 GAGCTAAGCTTATATGTTACAGAAAAATGACAGATGCGCTTCTTACGTATGATAGTG 1328
QY 1321 GAGAGATCGAGAGCCGCTGTGTGCTGCGGCTGCTGAGCTGGAGAGAGTGTATTCC 1380
DB 1329 GAGAGATCGAGAGCCGCTGTGTGCTGCGGCTGCTGAGCTGGAGAGAGTGTATTCC 1388
QY 1381 TGCCACATGCGCGGAGCGAGATGACACAACATGTCATCATCAATCTGTACAGTAAGTCAG 1440
DB 1389 TGCCACATGCGCGGAGCGAGATGACACAACATGTCATCATCAATCTGTACAGTAAGTCAG 1448
QY 1441 ATACACAGAACTTCTCTATACGTCAGTACTGGCCCGGCTCATGAGAGCCCGCTCT 1500
DB 1449 ATACACAGAACTTCTCTATACGTCAGTACTGGCCCGGCTCATGAGAGCCCGCTCT 1508
QY 1501 TCTACACTGTCCAGTACAGACTGGCTTGTGACTGTGCTGATTCACCGCTGGTCCATGCCA 1560
DB 1509 TCTACACTGTCCAGTACAGACTGGCTTGTGACTGTGCTGATTCACCGCTGGTCCATGCCA 1568
QY 1561 TGTGTGAGCCCACTTATGGAATCCACACAGATGCGCGGTGCAAGCGGAAATATCTCT 1620
DB 1569 TGTGTGAGCCCACTTATGGAATCCACACAGATGCGCGGTGCAAGCGGAAATATCTCT 1628
QY 1621 TCACTTCCAGGGGAGGAATGAGATGAGTCTGAGGTCATACCTTTCAGAGAGCCCGCTCT 1680
DB 1629 TCACTTCCAGGGGAGGAATGAGATGAGTCTGAGGTCATACCTTTCAGAGAGCCCGCTCT 1688
QY 1681 TCGAAGAGGAATGAGGGGAGCCCTCCGCGCACTACGATGACCGGATATGTCACACC 1740
DB 1689 TCGAAGAGGAATGAGGGGAGCCCTCCGCGCACTACGATGACCGGATATGTCACACC 1748
QY 1741 TGAAGCGGTGACGACAGCAAGCTGATCAGTCTGTGTGAATTCACCTGCAAAAAAC 1800
DB 1749 TGAAGCGGTGACGACAGCAAGCTGATCAGTCTGTGTGAATTCACCTGCAAAAAAC 1808
QY 1801 AGCCCAACCCAGCCGCGGAGCTGATGAGGCACTGTGTGAGAGCGGAGAGCCGCTGG 1860
DB 1809 AGCCCAACCCAGCCGCGGAGCTGATGAGGCACTGTGTGAGAGCGGAGAGCCGCTGG 1868
QY 1861 AATGTGTAAGCTTCCACCTTCCGCTCATCATTAACCCCGGGGAGCCCTGCTGGTTA 1920
DB 1869 AATGTGTAAGCTTCCACCTTCCGCTCATCATTAACCCCGGGGAGCCCTGCTGGTTA 1928
QY 1921 TTTCTCTGGGTGTCAACAGCGCTTTCGAAGCCCTGGAAGTCGGTCCGCTCCGGGTGG 1980
DB 1929 TTTCTCTGGGTGTCAACAGCGCTTTCGAAGCCCTGGAAGTCGGTCCGCTCCGGGTGG 1988

1981 TCGTGGGGAGGAGCTGACACTTCCCTACCAGACATGCTGCAGTGGAAACGAGCGGCC 2040
1989 TCGTGGGGAGGAGCTGACACTTCCCTACCAGACATGCTGCAGTGGAAACGAGCGGCC 2048
2041 TGGTGGTGGCAAGCTGCTGCTTACCAGAGTTCATTTCCGCTCAGAGAGCTTCCGATA 2100
2049 TGGTGGTGGCAAGCTGCTGCTTACCAGAGTTCATTTCCGCTCAGAGAGCTTCCGATA 2108
2101 GTGACCTCTGCTGATAGAGGCGGCAAGGCGCTTCTCTGGAGAGCTTACTTCTCCACTG 2160
2109 GTGACCTCTGCTGATAGAGGCGGCAAGGCGCTTCTCTGGAGAGCTTACTTCTCCACTG 2168
2161 CTGACAGTATTTTAAATTAATACCGTCTGCTGCTATGATTAAGACTGCGATCCAGATCCAGCCG 2220
2169 CTGACAGTATTTTAAATTAATACCGTCTGCTGCTATGATTAAGACTGCGATCCAGATCCAGCCG 2228
2221 CTCCCATCCGGGAGAGGCGGCGAGTGCATGCCACCGCTTACAGGCAAGGCGGCTGGAA 2280
2229 CTCCCATCCGGGAGAGGCGGCGAGTGCATGCCACCGCTTACAGGCAAGGCGGCTGGAA 2288
2281 CTGACCCCAACATGCTGACCAACGAGGAGCTGACCTGGGCGCAGTGAAGACGAGCCG 2340
2289 CTGACCCCAACATGCTGACCAACGAGGAGCTGACCTGGGCGCAGTGAAGACGAGCCG 2348
2341 CCGACGCTTACCAGATTAATCTCCGCAATTTCACTGCTGACTGACTGACTTTTACCGCA 2400
2349 CCGACGCTTACCAGATTAATCTCCGCAATTTCACTGCTGACTGACTGACTTTTACCGCA 2408
2401 GGTGGAAAGTGTGCTGAGGCGCTTCCATCTTTCCCGCAACCTGCCCTTGAACCTGTGT 2460
2409 GGTGGAAAGTGTGCTGAGGCGCTTCCATCTTTCCCGCAACCTGCCCTTGAACCTGTGT 2468
2461 TGGCTCAGAGGCGCAATTTCTGGGCTCAGGAGCTGCTGGGCGCTTATGGTGTGTGAG 2520
2469 TGGCTCAGAGGCGCAATTTCTGGGCTCAGGAGCTGCTGGGCGCTTATGGTGTGTGAG 2528
2521 CTGGGGGTTCTGGCAAGAAATTTCAAGCAGCGCTTGGAGCAATGTTCCCGAGACAGT 2580
2529 CTGGGGGTTCTGGCAAGAAATTTCAAGCAGCGCTTGGAGCAATGTTCCCGAGACAGT 2588
2581 TCGAGTGGGATGTTGACTTATGAGGAGGAGAGTGGCTTATGAATCTTTAGAGAGG 2640
2589 TCGAGTGGGATGTTGACTTATGAGGAGGAGAGTGGCTTATGAATCTTTAGAGAGG 2648
2641 TGAATGGCTTCCCTTACCTGAACAGAGTGTGTGTGTGAATTTCCCAAGCTCCAT 2700
2649 TGAATGGCTTCCCTTACCTGAACAGAGTGTGTGTGTGAATTTCCCAAGCTCCAT 2708
2701 CAGAGACCTTCTGTGGCTGACATTTGGGCTTCCATGATGTGTCCGTACTGAGAGA 2760
2709 CAGAGACCTTCTGTGGCTGACATTTGGGCTTCCCAATGATGTGTCCGTACTGAGAGA 2768
2761 ACAGTTTGAACAACGATTTCTTACCTGGAAATGAAATGAGAGAGGCGCATCTGTCCA 2820
2769 ACAGTTTGAACAACGATTTCTTACCTGGAAATGAAATGAGAGAGGCGCATCTGTCCA 2828
2821 TTGATAGCAGTGTCACTCCGCGCATGAGAAATCATGTTGGGTTCCGGGTGTGAGAG 2880
2829 TTGATAGCAGTGTCACTCCGCGCATGAGAAATCATGTTGGGTTCCGGGTGTGAGAG 2888
2881 AAGCTGGGAGCGGATGCTGGGCTTCCCTGGGCTTACCAAGATGGAGATCCCCATC 2940
2889 AAGCTGGGAGCGGATGCTGGGCTTCCCTGGGCTTACCAAGATGGAGATCCCCATC 2948
2941 AGTCTGGGCTTCACTCAACTCACTTCTGTGAGCTGTCAATGCTGAGACAGTGTG 3000
2949 AGTCTGGGCTTCACTCAACTCACTTCTGTGAGCTGTCAATGCTGAGACAGTGTG 3008
3001 CCTTCTTACAGATTAATGCTTACCTGATTTCTTATGTGATGAGCCAGGCGCATCGG 3060
3009 CCTTCTTACAGATTAATGCTTACCTGATTTCTTATGTGATGAGCCAGGCGCATCGG 3068

3061 ACATGGTGGATGAATACATCAACTGAGAGACATGGCATGAATCTTGTGTCCACA 3120
3069 ACATGGTGGATGAATACATCAACTGAGAGACATGGCATGAATCTTGTGTCCACA 3128
3121 TCACCTGGAGAGCCGCCCATCAAGTGTACCTCAAGGTGACATTTCCGATCCAGATGCC 3180
3129 TCACCTGGAGAGCCGCCCATCAAGTGTACCTCAAGGTGACATTTCCGATCCAGATGCC 3188
3181 CTGAGGCGCTGTCTCAATGATGACTCCACTTCCAGAGGCGGCAAGTGCATCAATCT 3240
3189 CTGAGGCGCTGTCTCAATGATGACTCCACTTCCAGAGGCGGCAAGTGCATCAATCT 3248
3241 TGGTGAAGTGTACGCTTCAATGCTTCTGTGTACAGCAGTTCAGGTGATTTGTGTC 3300
3249 TGGTGAAGTGTACGCTTCAATGCTTCTGTGTGTACAGCAGTTCAGGTGATTTGTGTC 3308
3301 TCTTCAAGACAGGCGTCCCATGACAAAGAGTTCAGATTCATCTAGGCGCAGC 3360
3309 TCTTCAAGACAGGCGTCCCATGACAAAGAGTTCAGATTCATCTAGGCGCAGC 3368
3361 GCACGCTCTGGGAGAGAGATGAGCAGAGGAGAGATGCTCCCAAGGTTCTTAGGCA 3420
3369 GCACGCTCTGGGAGAGAGATGAGCAGAGGAGAGATGCTCCCAAGGTTCTTAGGCA 3428
3421 TTGACAGACCTTGGGACATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3480
3429 TTGACAGACCTTGGGACATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3488
3481 CAGAGAGTGTGAGAGAAACCGCTGCTTATCTTGAATGACGACACCTGGGCGCTGGAG 3540
3489 CAGAGAGTGTGAGAGAAACCGCTGCTTATCTTGAATGACGACACCTGGGCGCTGGAG 3548
3541 CCCTGGGCGAGTCCCGGGGTTCCACACAGGCGACTGACTGATAGCTTACACTGAG 3600
3549 CCCTGGGCGAGTCCCGGGGTTCCACACAGGCGACTGACTGATAGCTTACACTGAG 3608
3601 ACTGTGGCAGCTGTGAGAGTCACTACACCGTCTACGCCAGAGCAGCTGTGTG 3660
3609 ACTGTGGCAGCTGTGAGAGTCACTACACCGTCTACGCCAGAGCAGCTGTGTG 3668
3661 GTTTTACATTCATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3720
3669 GTTTTACATTCATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3728
3721 ATTCAAGGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3780
3729 ATTCAAGGCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3788
3781 TATATATTTTGGCTGGGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3840
3789 TATATATTTTGGCTGGGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3848
3841 TCCCTTGTCTGATCTGAGGAGATGGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3900
3849 TCCCTTGTCTGATCTGAGGAGATGGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3908
3901 AAGATCATCTCCATCTCCAGCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3960
3909 AAGATCATCTCCATCTCCAGCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3968
3961 TGGGCGCTTGGGCGCTGGAGAGGCTTCTTCCATACCCTTCTGTGTGTGTGTGTGTGTGT 4020
3969 TGGGCGCTTGGGCGCTGGAGAGGCTTCTTCCATACCCTTCTGTGTGTGTGTGTGTGTGT 4028
4021 CCTGTAACTACATTTCTTCCAGTGAATCCCTGGAGAGGCTGACCTGTGTGTGTGTGTGT 4080
4029 CCTGTAACTACATTTCTTCCAGTGAATCCCTGGAGAGGCTGACCTGTGTGTGTGTGTGT 4088
4081 TCAAGCTTCTGT 4140
4089 TCAAGCTTCTGT 4148
4141 TCCGTACTTATCCGCTCTCCCATTTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4200

Db	5229	TCATAGTTGCCCTTACCTGACCTCCATCTGTATCTCTTTTACAGTAGCCTTCTTCCT	5288
QY	5281	CGGGGAGCCAAAGATGTGGTGTGTGGCGCTATATTGTGGCTGCTPAATTCATGTGGTTTC	5340
Db	5289	CGGGGAGCCAAAGATGTGGTGTGTGGCGCTATATTGTGGCTCTATTTCATGTGGTTTC	5348
QY	5341	TTTTATGTGAGAACTCATATCTGACTTTCAGTGGACTCGGGTGAAGCCGGGCGCTCTG	5400
Db	5349	TTTTAATGTGAGAACTCATATCTGACTTTCAGTGGACTCGGGTGAAGCCGGGCGCTCTG	5408
QY	5401	TGTGGGGAGACCCCTTTTGGCGGAGTCAGTAGCTGGGGCGCTCTGTGTGTGAAGCCA	5460
Db	5409	TGTGGGGAGACCCCTTTTGGCGGAGTCAGTAGCTGGGGCGCTCTGTGTGTGAAGCCA	5468
QY	5461	GGGCGCTTCCTTTAGTGGAGCCAGGTTTGTGGGCGCCCAATGTGCTAGCTGGTGGATCTAG	5520
Db	5469	GGGCGCTTCCTTTAGTGGAGCCAGGTTTGTGGGCGCCCAATGTGCTAGCTGGTGGATCTAG	5528
QY	5521	AAGGCTGAGTGTGTGACACCAAAACATGCCGAGGAGGGGCTGTGTGCCGCTTC	5580
Db	5529	AAGGCTGAGTGTGTGACACCAAAACATGCCGAGGAGGGGCTGTGTGCCGCTTC	5588
QY	5581	CAACAGGAGACAGCCCTCTCTGACCCCTGAAAGAAACACTGCTTGAAGACATGACAGACAG	5640
Db	5589	CAACAGGAGACAGCCCTCTCTGACCCCTGAAAGAAACACTGCTTGAAGACATGACAGACAG	5648
QY	5641	CTGTAGGGGAGCAGCGCTCTCAGCGAAGAGCAGCAAGTGGCCCACTGTACATGAGCA	5700
Db	5649	CTGTAGGGGAGCAGCGCTCTCAGCGAAGAGCAGCAAGTGGCCCACTGTACATGAGCA	5708
QY	5701	GGTCTTCTCACCACGGGAAAGCCGCGCACTGTGACTCGCTTGAATGGGAAAGCGGCG	5760
Db	5709	GGTCTTCTCACCACGGGAAAGCCGCGCACTGTGACTCGCTTGAATGGGAAAGCGGCG	5768
QY	5761	CCACAGACCCCGGGTCTCTGTTGCTGTGTGGGCGCCCTGTGGCCACCTTGTCTTGACT	5820
Db	5769	CCACAGACCCCGGGTCTCTGTTGCTGTGTGGGCGCCCTGTGGCCACCTTGTCTTGACT	5828
QY	5821	CGCAGGGTGCAGAGAGCGCTCGTTCCTGTGGGTGGCGGGTTCGTCGCTCCGTTTGGGCTG	5880
Db	5829	CGCAGGGTGCAGAGAGCGCTCGTTCCTGTGGGTGGCGGGTTCGTCGCTCCGTTTGGGCTG	5888
QY	5881	TCTTTACCATTAACACCGTCCAGGGCTCTGCAGGCACTGTGAGCGCTGCTCCCTGGGCA	5940
Db	5889	TCTTTACCATTAACACCGTCCAGGGCTCTGCAGGCACTGTGAGCGCTGCTCCCTGGGCA	5948
QY	5941	GTCCTCTCTCCGTGTGAGACTGTGCTCAGGCAAGGCTCACCAGCTGGGGCTCTGTCCGGA	6000
Db	5949	GTCCTCTCTCCGTGTGAGACTGTGCTCAGGCAAGGCTCACCAGCTGGGGCTCTGTCCGGA	6008
QY	6001	AGGATGGAGTCTTTTGGGAGCTGGGCGCGGACAGACTGGGGAGACTTCTTGTGTGGGG	6060
Db	6009	AGGAT--GGAATCTTTCTGGGAGCTGGGCGCGGACAGACTGGGGAGACTTCTTGTGTGGGG	6067
QY	6061	GAAAGCTTGATATTCATGCCAGTCCATCCACCCACCCCTTTTGTGTCAGAGCAACAAG	6120
Db	6068	GAAAGCTTGATATTCATGCCAGTCCATCCACCCACCCCTTTTGTGTCAGAGCAACAAG	6127
QY	6121	GTCTTACATGGATTTTTGT 6140	
Db	6128	GTCTTACATGGATTTTTGT 6147	

RESULT	5
AF029231	
LOCUS	5854 bp mRNA linear
DEFINITION	Homo sapiens EXT homolog mRNA, complete cds.
ACCESSION	AF029231
VERSION	AF029231.1 GI:4103863
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 5854)
Deng, H.-X., Xu, L., Xia, J.-H., Fan, C., Pan, Q., Liu, C.-Y. and
Ruan, Q.-G.
TITLE
Molecular cloning of a candidate gene for multiple exostoses
JOURNAL
Unpublished
2 (bases 1 to 5854)
Deng, H.-X., Xu, L., Xia, J.-H., Fan, C., Pan, Q., Liu, C.-Y. and
Ruan, Q.-G.
TITLE
Direct Submision
JOURNAL
Submitted (07-OCT-1997) National Laboratory of Medical Genetics,
Human Medical University, 88 Xiangya Road, Changsha, Hunan 410078,
P.R. of China

FEATURES
source
Location/Qualifiers
1..5854
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
263..3022
/note="candidate gene for multiple exostoses"
/product="EXT homolog"
/protein_id="AAD01877.1"
/db_xref="GI:4103884"
/translation="MTGYTMLRNGANGNGGTGCMTRPMSNRITLTLSTFLVLYLFFP
LIAHYLTTLDEADAGKRITGPRVNGNECEVHLDLCRIRESVSEELQLEAKRQ
LNSLAKNLKIEACKSRLENKODLKNVISOHESYKLMONOPKLSLRRL
PEKADAGLPPEPKATGRCRLHNCFSRCPLTSGFVYVDSOEFSGYSLDPLVQAF
QATARAANYVTEENADIACLYLVGEMOEPVLYRALEKOLKSLPHNMTDHNHVI
NLSRSDPNLLYNSGTGRAMYASOSTFYVORPDLVSLVLSVAMSMEIPLPO
VFKRKYLTFTGEGEKIESLRSLQBARSEEMEDPPADYDRIATILAKAVDSKLD
OVLVEFTCKNOPKPSLPTEWALCGEREDRLKLTSLALITTPDPRILVISGCAIR
LFEALVGAIVPVLEOVQLPYQDMLQNMENALVYKPRVTEVHFLRLSLSDDLAM
RROGRLMETEYSTADISFNTVLAIRTRIQIPARIEEALAEIPIHSGKAQTDLP
MADNDLDLGGPVEETPPYASPRYLRFNFTLVTDYFSNNKCAAGPHLPFPDPDVL
SPAKFGSGTGRPRTPGGAGSGSKRFOALAGNVBPDPYVMLTYEPREVLNLSER
LNGLPYLNKYVYVWNSPKLPSDLPLPDQGVIVYVTRKNSLNRLPRMNIETBEAI
LSIDDAHLRHEIMFGFRVWRKARDRLVGPGRYHAMDIPHQSLYNVNSCELSMV
LTGAALFHKYLYLYSYVPOALRDMVDEYINCEDIANMFLVSHTRKPRIVTSRMT
FRPCPCPOLSHDSDHFHERHKINFVKVYGMPLLYTQFRVDSVLEFTRILPHDKTK
CFKFI"

BASE COUNT 1204 a 1612 c 1610 g 1428 t
ORIGIN

Query Match 94.5%; Score 5834.6; DB 9; Length 5854;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 5837; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

332 GGCACCTTTTATCGTTTGATTAAGACATGTTCTTTGGTCAACAGCCAGAACTT 391
|||||
Db 1 GGCACCTTTTATCGTTTGATTAAGACATGTTCTTTGGTCAACAGCCAGAACTT 60
392 AAAATGCGGGAATAGGGGAGAGACCAATTCAGTGCAGCGAGAGAAATGAATGTT 451
Db 61 AAAATGCGGGAATAGGGGAGAGACCAATTCAGTGCAGCGAGAGAAATGAATGTT 120
452 CATTTATTTGGTGCCTTGTCTGGGAGACACTAATCTTCTTGAAGAGTGTCACTGAA 511
|||||
Db 121 CATTTATTTGGTGCCTTGTCTGGGAGACACTAATCTTCTTGAAGAGTGTCACTGAA 180
512 ACAGAGATGCTTTTGTGGAATAGCAACCAATGTTATGGCAGTGAACCCGACGTGATCTG 571
Db 181 ACAGAGATGCTTTTGTGGAATAGCAACCAATGTTATGGCAGTGAACCCGACGTGATCTG 240
572 GGGGGAGGCTGCAGAGAGACTCATGACAGCTATACCATGCTCCGGAATGGGGCCCGGG 631
Db 241 GGGGGAGGCTGCAGAGAGACTCATGACAGCTATACCATGCTCCGGAATGGGGCCCGGG 300
632 GAACGAGGTGACAGACTGATGCTGTGATGTCACACCCGATCCGCTCAGTGGCTCAG 691
|||||
Db 301 GAACGAGGTGACAGACTGATGCTGTGATGTCACACCCGATCCGCTCAGTGGCTCAG 360

Qy 692 CTTACAGCTCTTTGTGTCATCTGCTCTTCCCGCTCATCCGCCACTATTACCTCAACC 751
|||||
Db 361 CTTACAGCTCTTTGTGTCATCTGCTCTTCCCGCTCATCCGCCACTATTACCTCAACC 420
752 TCTGAGTGAAGGCGATGAGGAGGCAAGGCAATTTTGGTCCCGGGTGGGGAAGAGCT 811
|||||
Db 421 TCTGAGTGAAGGCGATGAGGAGGCAAGGCAATTTTGGTCCCGGGTGGGGAAGAGCT 480
812 GTGCGAGGTGAAGCAAGCTGCTGGATCTGTGCCGATCCGGAGTGGGTAGTGAAGAGCT 871
|||||
Db 481 GTGCGAGGTGAAGCAAGCTGCTGGATCTGTGCCGATCCGGAGTGGGTAGTGAAGAGCT 540
872 CCTGACGTTGAGAGCCAGCCAGCCAGAGCTGAACAGAGATCCGACGTAATCTGAA 931
Qy 872 CCTGACGTTGAGAGCCAGCCAGCCAGAGCTGAACAGAGATCCGACGTAATCTGAA 931
|||||
Db 541 CCTGACGTTGAGAGCCAGCCAGCCAGAGCTGAACAGAGATCCGACGTAATCTGAA 600
932 GATCGAAGCTGTGAAGAAGCAATTGAGAACGCCAAGCAGAGACCTGCTCAGCTCAAGAA 991
Db 601 GATCGAAGCTGTGAAGAAGCAATTGAGAACGCCAAGCAGAGACCTGCTCAGCTCAAGAA 660
992 TGTCAATCAGCCAGAGCCAGACATTCCTACAAAGAGCTCATGGCCAGAACCCAGAGCT 1051
Db 661 TGTCAATCAGCCAGAGCCAGACATTCCTACAAAGAGCTCATGGCCAGAACCCAGAGCT 720
1052 GTCCCTGCCCATCCGACTGCTGCCAGAGAAAGACATGCCGGCTCCCTCCCGCAAGGC 1111
Qy 1052 GTCCCTGCCCATCCGACTGCTGCCAGAGAAAGACATGCCGGCTCCCTCCCGCAAGGC 1111
721 GTCCCTGCCCATCCGACTGCTGCCAGAGAAAGACATGCCGGCTCCCTCCCGCAAGGC 780
1112 CACTCGGGGCTGGCGGCTGCACAACTGCTTGTATTTCTGCTGTTGGCCTCTCAGCTCTGG 1171
|||||
Db 781 CACTCGGGGCTGGCGGCTGCACAACTGCTTGTATTTCTGCTGTTGGCCTCTCAGCTCTGG 840
1172 CTTCCCGGCTTACGTTATGACAGTGAACAGATTGTTCTTTGGCAGCTTACCTGATCCCTT 1231
841 CTTCCCGGCTTACGTTATGACAGTGAACAGATTGTTCTTTGGCAGCTTACCTGATCCCTT 900
1232 GGTCAAGAGAGCTTTTACAGCCGACAGACAGAGTAACTGTTATGTTACAGAAATGCAAA 1291
Db 901 GGTCAAGAGAGCTTTTACAGCCGACAGACAGAGTAACTGTTATGTTACAGAAATGCAAA 960
1292 CATCGCCTGCTTACGTTATGATAGTGGAGAGATGACAGAGCCCGGTGGTGGCGGC 1351
Db 961 CATCGCCTGCTTACGTTATGATAGTGGAGAGATGACAGAGCCCGGTGGTGGCGGC 1020
1352 TGTGAGCTGGAGAGAGCACTTGTATTCCTGCGACACTGGCGGAGATGAGACACAACA 1411
Db 1021 TGTGAGCTGGAGAGAGCACTTGTATTCCTGCGACACTGGCGGAGATGAGACACAACA 1080
1412 TGTCAATCATCATCTGTACAGTGAAGTACAGATACAGAGAACTTCTCTATACGTCACTG 1471
Db 1081 TGTCAATCATCATCTGTACAGTGAAGTACAGATACAGAGAACTTCTCTATACGTCACTG 1140
1472 TGGCGCTGCATGCTGGTGGCCAGTCCACCTTCTACACTGTCTCAGTACAGACTGGCTTTGA 1531
Db 1141 TGGCGCTGCATGCTGGTGGCCAGTCCACCTTCTACACTGTCTCAGTACAGACTGGCTTTGA 1200
1532 CTTGATCGTATCACCGCTGCTCATGCTCATGCTGAGGCCCAACTTATGAAATCCACAC 1591
Db 1201 CTTGATCGTATCACCGCTGCTCATGCTCATGCTGAGGCCCAACTTATGAAATCCACAC 1260
1592 ACAGGTGCGGGTGAAGCGGAATATCTCTTACACTTCCAGGGGAGAGAATGATGATCTCT 1651
Db 1261 ACAGGTGCGGGTGAAGCGGAATATCTCTTACACTTCCAGGGGAGAGAATGATGATCTCT 1320
1652 GAGGTGATGCTTACAGAGAGCCGCTCTTGAAGAGAAATGAGGGGAGACCTCCCGG 1711
Db 1321 GAGGTGATGCTTACAGAGAGCCGCTCTTGAAGAGAAATGAGGGGAGACCTCCCGG 1380
1712 CGACTAGATGACCGGATCATTCGCCAACCCTGAAAGGGGTGACAGAGCAAGCTGATGATCA 1771
Db 1381 CGACTAGATGACCGGATCATTCGCCAACCCTGAAAGGGGTGACAGAGCAAGCTGATGATCA 1440
1772 GGTCTGTGTGAATTCACCTGCAAAAACCAAGCCCAAGCCAGCTGCGCATGATGGGC 1831

Db 1441 GGTCTGTTGGATTTACCTGCAAAAAACACCCAGCCGACCTGACACTGATGGGC 1500
QY 1832 ACTGTGGAGAGCGGGAGAGACCGCTTGGAAATGTGTGAAGTCTCCACCTTCCCTCAT 1891
Db 1501 ACTGTGGAGAGCGGGAGAGACCGCTTGGAAATGTGTGAAGTCTCCACCTTCCCTCAT 1560
QY 1892 CATTTACCCCCGGGGACCCCTGCTTGTATTTCTCTGAGGTGTCAACAGAGGCTTCGA 1951
Db 1561 CATTTACCCCCGGGGACCCCTGCTTGTATTTCTCTGAGGTGTCAACAGAGGCTTCGA 1620
QY 1952 AGCCCTTGAAGTGTGGTCCCTCCCGGTGTGTGTGGGGAGAGTCCAGCTTCCCTACA 2011
Db 1621 AGCCCTTGAAGTGTGGTCCCTCCCGGTGTGTGTGGGGAGAGTCCAGCTTCCCTACA 1680
QY 2012 GGCATGTGTGACATGAGACAGAGCGGCCCTGGTGTGGCCAAACCTCTGTTCACAGAT 2071
Db 1681 GGCATGTGTGACATGAGACAGAGCGGCCCTGGTGTGGCCAAACCTCTGTTCACAGAT 1740
QY 2072 TCAATTTCCCTCAGAACGCTTCCGATAGTACCTCTGAGCTATGAGGCGGCAAGGCG 2131
Db 1741 TCAATTTCCCTCAGAACGCTTCCGATAGTACCTCTGAGCTATGAGGCGGCAAGGCG 1800
QY 2132 CTTTCTCTGGGAGACTTACTTCTCCACTGTCTACAGTATTTTAAATACCTGTGCTAT 2191
Db 1801 CTTTCTCTGGGAGACTTACTTCTCCACTGTCTACAGTATTTTAAATACCTGTGCTAT 1860
QY 2192 GATTAGACATGACATCCAGATCCACGCGCTCCCATCCGGGAGAGAGCGGCGACTAGAT 2251
Db 1861 GATTAGACATGACATCCAGATCCACGCGCTCCCATCCGGGAGAGAGCGGCGACTAGAT 1920
QY 2252 CCCCCACGCTTCAGGACAGGCGGCTGGAATGACCCCAACATGAGTGAACAACGGGAGACT 2311
Db 1921 CCCCCACGCTTCAGGACAGGCGGCTGGAATGACCCCAACATGAGTGAACAACGGGAGACT 1980
QY 2312 GGACCTGGGGCCAGGAGGAGGAGCGGCCCTACAGCTCACCCAGATACCTCCGCAATTT 2371
Db 1981 GGACCTGGGGCCAGGAGGAGGAGCGGCCCTACAGCTCACCCAGATACCTCCGCAATTT 2040
QY 2372 CACTCTGACTGTCACTACTTTTACCGCAGCTGGAACGTGTCTCCAGAGGCTTTCACATCT 2431
Db 2041 CACTCTGACTGTCACTACTTTTACCGCAGCTGGAACGTGTCTCCAGAGGCTTTCACATCT 2100
QY 2432 TTTTCCCCACACATCCCTTTTAAACCTGTGTGCCCTCAGAGGCCAAATTTCTTGGGCTCAG 2491
Db 2101 TTTTCCCCACACATCCCTTTTAAACCTGTGTGCCCTCAGAGGCCAAATTTCTTGGGCTCAG 2160
QY 2492 GACTGGCTTTCGCGCTATTGTGTGTGAGCTGGGGCTTTCGCAAGGAATTTTCAGCGAC 2551
Db 2161 GACTGGCTTTCGCGCTATTGTGTGTGAGCTGGGGCTTTCGCAAGGAATTTTCAGCGAC 2220
Y 2552 GCTTGGAGGCAATGTTTCCCGAGAGCAGTTACAGGTGTGTGATGTGTACTATAGCGGGA 2611
Db 2221 GCTTGGAGGCAATGTTTCCCGAGAGCAGTTACAGGTGTGTGATGTGTACTATAGCGGGA 2280
QY 2612 GGAAGGCTTATGAACTCTTTAGAGAGCTGAATGGCTCCCTTACTGTAACAAGGTGCT 2671
Db 2281 GGAAGGCTTATGAACTCTTTAGAGAGCTGAATGGCTCCCTTACTGTAACAAGGTGCT 2340
QY 2672 GGTGTGTGGAATTTCCCAACCTGCAATCAGAGACCTTCTGTGGCTGACATTTGGCGT 2731
Db 2341 GGTGTGTGGAATTTCCCAACCTGCAATCAGAGACCTTCTGTGGCTGACATTTGGCGT 2400
QY 2732 TTCCATCATGTGTGTCTCTACTGAGAGAACAGTTTGAACAACGATTTCTTACCTGGAA 2791
Db 2401 CCCCATCATGTGTGTCTCTACTGAGAGAACAGTTTGAACAACGATTTCTTACCTGGAA 2460
QY 2792 TGAATGAGACAGAGGCAATCTGTCTCATTTGATGACATGCTCATCCGCAATGAGGA 2851
Db 2461 TGAATGAGACAGAGGCAATCTGTCTCATTTGATGACATGCTCATCCGCAATGAGGA 2520
QY 2852 AATCATTTTGGGTTCGGGGTGTGAGAGAGCTCGGAGCCGATCGTGGGCTTCCCTGG 2911
|||||

Db 2521 AATCATTTTGGGTTCGGGGTGTGAGAGAGCTCGGAGCCGATCGTGGGCTTCCCTGG 2580
QY 2912 CCGTTACACGATGGGAGCATCCCCCATCACTGCTGTGCTTACAACTCCAACTACTCTG 2971
Db 2581 CCGTTACACGATGGGAGCATCCCCCATCACTGCTGTGCTTACAACTCCAACTACTCTG 2640
QY 2972 TGAGCTTCCATGTGTGTGACAGGTGTGCTGCTTCCCAAGATATATGCTTACCTGTA 3031
Db 2641 TGAGCTTCCATGTGTGTGACAGGTGTGCTGCTTCCCAAGATATATGCTTACCTGTA 2700
QY 3032 TTTCTATGTATGATGCCAGGCCATCCGGAGATGTGTGATGAAATACATCACTGTAAGA 3091
Db 2701 TTTCTATGTATGATGCCAGGCCATCCGGAGATGTGTGATGAAATACATCACTGTAAGA 2760
QY 3092 CATTTGCATGAATCTTCTTCTTCCCATCACTCTGGAAGCCCCCATCAAGGTGACCTC 3151
Db 2761 CATTTGCATGAATCTTCTTCTTCCCATCACTCTGGAAGCCCCCATCAAGGTGACCTC 2820
QY 3152 ACGGTGGACATTCGATGCGCCAGAGATGCCCTCAGAGGCTGTCTCATGATGACTCCACTT 3211
Db 2821 ACGGTGGACATTCGATGCGCCAGAGATGCCCTCAGAGGCTGTCTCATGATGACTCCACTT 2880
QY 3212 CCACGAGCGGCAACAGTGCATCACTTCTGCTGAAAGGTGTACGGCTTACATGCCCTCT 3271
Db 2881 CCACGAGCGGCAACAGTGCATCACTTCTGCTGAAAGGTGTACGGCTTACATGCCCTCT 2940
QY 3272 GTACACGAGTTCAGAGGTGTGCTGTCTCTTCAAGACACGCTTCCCATGACAAAC 3331
Db 2941 GTACACGAGTTCAGAGGTGTGCTGTCTCTTCAAGACACGCTTCCCATGACAAAC 3000
QY 3332 CAAGTCTTCAAGTTCATCTAGAGGGGACGACGAGGCTCTGGGGAAGAGATGACAGAGG 3391
Db 3001 CAAGTCTTCAAGTTCATCTAGAGGGGACGACGAGGCTCTGGGGAAGAGATGACAGAGG 3060
QY 3392 AGGAAGATGTCTCCCAAGGTTCTTAAAGCTTGCAGAGACCTTGGGCACTGTCTGTGG 3451
Db 3061 AGGAAGATGTCTCCCAAGGTTCTTAAAGCTTGCAGAGACCTTGGGCACTGTCTGTGG 3120
QY 3452 TGGCCAGAGGCTTGTGCGGGAAGGGGACGACGAGAGAGTGTGAAGAAACGCTTCTTAA 3511
Db 3121 TGGCCAGAGGCTTGTGCGGGAAGGGGACGACGAGAGAGTGTGAAGAAACGCTTCTTAA 3180
QY 3512 TCTTGAAGTACGACACTGGGCTGTGAGGCCCTGGGGAGTCCCGGGTTCGCCACAC 3571
Db 3181 TCTTGAAGTACGACACTGGGCTGTGAGGCCCTGGGGAGTCCCGGGTTCGCCACAC 3240
QY 3572 AGGCACTGACTATGACTTACACTGAGACTGTGGCACTGTGACAGATCACTACAC 3631
Db 3241 AGGCACTGACTATGACTTACACTGAGACTGTGGCACTGTGACAGATCACTACAC 3300
QY 3632 GTTGTACGCGCCAGGACAGCTGTGTGTCTTTCATTCAATTAACAATATTATGATT 3691
Db 3301 GTTGTACGCGCCAGGACAGCTGTGTGTCTTTCATTCAATTAACAATATTATGATT 3360
QY 3692 ATTTAAAAAGAAAGTTTCAAGTTTGCATTCAGGCTTATTTATATATGTGTGT 3751
Db 3361 ATTTAAAAAGAAAGTTTCAAGTTTGCATTCAGGCTTATTTATATATGTGTGT 3420
QY 3752 ATATATAATCATGACACACTTGCATATATATATTTTGGCTGGGGAGTGTAGTT 3811
Db 3421 ATATATAATCATGACACACTTGCATATATATATTTTGGCTGGGGAGTGTAGTT 3480
QY 3812 TGCCTTCTAAGGGGAGGAGCGCGAGGCTCTTGTGTGTGTGTGTGTGTGTGTGTGT 3871
Db 3481 TGCCTTCTAAGGGGAGGAGCGCGAGGCTCTTGTGTGTGTGTGTGTGTGTGTGTGT 3540
QY 3872 CCGGCTTGTGTGACTGTCTTATCTTAAAGATCATCTCCATCTTCCAGCGCATC 3931
Db 3541 CCGGCTTGTGTGACTGTCTTATCTTAAAGATCATCTCCATCTTCCAGCGCATC 3600
QY 3932 TGTGTGACGACACAGAAAGGATGAATGTGCGCTTGTGGGCGCTGTGACAAAGTCTCT 3991
Db 3601 TGTGTGACGACACAGAAAGGATGAATGTGCGCTTGTGGGCGCTGTGACAAAGTCTCT 3660
|||||

OY	3992	TCCTTACCCTTTCTGTTCSCAGTCACAACTGTAACTCACATTCTTCTCCCATGATC	4051
Db	3661	TCCTTACCCTTTCTGTTCSCAGTCACAACTGTAACTCACATTCTTCTCCCATGATC	3720
OY	4052	CCCTGGAGCCCTGACCCCTGTGTGGCTGTTCAGCTTCCGCGCTGGGGCCAGGAAATTT	4111
Db	3721	CCCTGGAGCCCTGACCCCTGTGTGGCTGTTCAGCTTCCGCGCTGGGGCCAGGAAATTT	3780
OY	4112	TGAGGATTTATCTTTAGGCCAGGCTTGCCCTCGACTTAATCCCTCTCTCCATTTCTCT	4171
Db	3781	TGAGGATTTATCTTTAGGCCAGGCTTGCCCTCGACTTAATCCCTCTCTCCATTTCTCT	3840
OY	4172	CTTCTTTGACAGAAATGATGAGGAACAAABAGTGAAGAAATAGGGCTGMAAGCCAC	4231
Db	3841	CTTCTTTGACAGAAATGATGAGGAACAAABAGTGAAGAAATAGGGCTGMAAGCCAC	3900
OY	4232	TCCAGATGGCTCTTTCTATCTGCTCTTCTGTGAAACACAGCTGTGGGCTCAG	4291
Db	3901	TCCAGATGGCTCTTTCTATCTGCTCTTCTGTGAAACACAGCTGTGGGCTCAG	3960
OY	4292	CGTTTCTGAAGTCTCTTTCTTGATTTGACAGAGATCAGCAGCTGCATCTGCTGT	4351
Db	3961	CGTTTCTGAAGTCTCTTTCTTGATTTGACAGAGATCAGCAGCTGCATCTGCTGT	4020
OY	4352	GGCTGGAAGTGTTTTGAGGTCAGGCTCCCTCCCTAGTGTAGACAAAGCCAGTGTCTT	4411
Db	4021	GGCTGGAAGTGTTTTGAGGTCAGGCTCCCTCCCTAGTGTAGACAAAGCCAGTGTCTT	4080
OY	4412	CGAGGAACCCACCCGCGCTGGCGGGAAATTTACAGCAAGCGCCCTGGCTTGGGTAATT	4471
Db	4081	CGAGGAACCCACCCGCGCTGGCGGGAAATTTACAGCAAGCGCCCTGGCTTGGGTAATT	4140
OY	4472	CCTTGTGTAAATTCACCTTCCCGCGCTGTGTGAGCCCCATCTGTGTATCTGTG	4531
Db	4141	CCTTGTGTAAATTCACCTTCCCGCGCTGTGTGAGCCCCATCTGTGTATCTGTG	4200
OY	4532	GTTTTTGGACCCCTTAATGTCAAGCTTGGCTGTAGAGATCCCGGAGTTTGGTATGTCTAG	4591
Db	4201	GTTTTTGGACCCCTTAATGTCAAGCTTGGCTGTAGAGATCCCGGAGTTTGGTATGTCTAG	4260
OY	4592	AACAATGGAGGCTGTGATTTGCTGTGTAGCTCACATCCAGCCCTTGGAACTTAACGGGC	4651
Db	4261	AACAATGGAGGCTGTGATTTGCTGTGTAGCTCACATCCAGCCCTTGGAACTTAACGGGC	4320
OY	4652	ATTTCACAACCCGAGTTACACATTTCCACATCCCTGCTTAAGGATTTCTTCCCTGGGCTAA	4711
Db	4321	ATTTCACAACCCGAGTTACACATTTCCACATCCCTGCTTAAGGATTTCTTCCCTGGGCTAA	4380
OY	4712	ACTGAAATTAAGCTTAATTTTGGGTACAGGTGACAGTAGGGGAACCTTAGAGAGGTGTAG	4771
Db	4381	ACTGAAATTAAGCTTAATTTTGGGTACAGGTGACAGTAGGGGAACCTTAGAGAGGTGTAG	4440
OY	4772	TGGCAATTTGTCAGGGAATTTAGCCATGAGCTGTTCCTTGAACCCACTTCTGGAAGTGG	4831
Db	4441	TGGCAATTTGTCAGGGAATTTAGCCATGAGCTGTTCCTTGAACCCACTTCTGGAAGTGG	4500
OY	4832	AGTTGACTCTGGAAGTTTCTACCAACTGAACAAAGCTCAGGTTTGTCTGTGTAATGCA	4891
Db	4501	AGTTGACTCTGGAAGTTTCTACCAACTGAACAAAGCTCAGGTTTGTCTGTGTAATGCA	4560
OY	4892	CATGCGCTTAAGCCAGTTCCGCTTCCCTAGACCTTGGACATCTGTGCTTATTTCTTGG	4951
Db	4561	CATGCGCTTAAGCCAGTTCCGCTTCCCTAGACCTTGGACATCTGTGCTTATTTCTTGG	4620
OY	4952	AATAGCTTCTCTGTGACTGCTGTACAGCTGTACACAGTGGTCTCTTCAAGTACTGTTTGAAG	5011
Db	4621	AATAGCTTCTCTGTGACTGCTGTACAGCTGTACACAGTGGTCTCTTCAAGTACTGTTTGAAG	4680
OY	5012	CTGGGCTCTTTTGTGTAGCTTCCACCCACCTGTAGGGCTAGCTCGGCTTAAAGGAACCTCT	5071
Db	4681	CTGGGCTCTTTTGTGTAGCTTCCACCCACCTGTAGGGCTAGCTCGGCTTAAAGGAACCTCT	4740

0Y	5072	CCCATTTGGCAAAACGGACACCCGGGCGCGGACGAGACTGTTTTCCAAAGGTTTCCCGGCC	5131
Db	4741	CCCCATTTGGCAAAACGGACACCCGGGCGCGGACGAGACTGTTTTCCAAAGGTTTCCCGGCC	4800
0Y	5132	CCCAACCCACGATCAGGCTTACGCTCCCGTGTGAGGCAAGTGTGGTTATGTTCCACAG	5191
Db	4801	CCAAACCCACGATCAGGCTTACGCTCCCGTGTGAGGCAAGTGTGGTTATGTTCCACAG	4860
0Y	5192	TGGGGGTCAGAGCGCCCTCTCAGAACTTTCAGTTGGCCCTCTACCTGACTGCTG	5251
Db	4861	TGGGGGTCAGAGCGCCCTCTCAGAACTTTCAGTTGGCCCTCTACCTGACTGCTGACTG	4920
0Y	5252	TATTCCTTTTAGCAGTAGGCTTCTTCCCTCGGGGAGCCAAAGCTGTGTGTGGGGCT	5311
Db	4921	TATTCCTTTTAGCAGTAGGCTTCTTCCCTCGGGGAGCCAAAGAGTGTGTGTGGGGCT	4980
0Y	5312	ATATTGGGCGCATATTCATCTGTTTTCTTTTAATGTAGAACTCAGATACAGACTTC	5371
Db	4981	ATATTGGGCGCATATTCATCTGTTTTCTTTTAATGTAGAACTCAGATACAGACTTC	5040
0Y	5372	AGTGGGACTGCTGAGCGCGGGGCGCTGTGTGTGTGGGACCCCTTTAGCGGACTCAGT	5431
Db	5041	AGTGGGACTGCTGAGCGCGGGGCGCTGTGTGTGTGGGACCCCTTTAGCGGACTCAGT	5100
0Y	5432	GAGCTGGGGCGCTGTGTGTGTGTGGAGCCAGAGGCTCTCCCTTTACTGGAGCAGGTTGTCT	5491
Db	5101	GAGCTGGGGCGCTGTGTGTGTGTGGAGCCAGAGGCTCTCCCTTTACTGGAGCAGGTTGTCT	5160
0Y	5492	GGGGCCCGAATGTCACTGAGTGGATCTTAAAGAGGGCTGAGTGTGTGACACCAAAACATGC	5551
Db	5161	GGGGCCCGAATGTCACTGAGTGGATCTTAAAGAGGGCTGAGTGTGTGACACCAAAACATGC	5220
0Y	5552	CGCAGGGAGGGCTGTGTGTGCCGGTGTCTCCAAAGAGACAGCCCTCTTGACCCCTGAAG	5611
Db	5221	CGCAGGGAGGGCTGTGTGTGCCGGTGTCTCCAAAGAGACAGCCCTCTTGACCCCTGAAG	5280
0Y	5612	GAAACATGTGCTTAAAGATCTGCAGACAGGCTCTGAGGGGACAGCCCTCTCAGGAGAGG	5671
Db	5281	GAAACATGTGCTTAAAGATCTGCAGACAGGCTCTGAGGGGACAGCCCTCTCAGGAGAGG	5340
0Y	5672	CAGCAAGGTGGCCACAGTGTCACTGGGTGAGAGTGTCTTCAACAGGAGGAAAGCCGCGCAC	5731
Db	5341	CAGCAAGGTGGCCACAGTGTCACTGGGTGAGAGTGTCTTCAACAGGAGGAAAGCCGCGCAC	5400
0Y	5732	TGTGACTGTGATGAGATGGGAAAGCGGCGCACAGACCCCGGGTCTCGTTGGCTGTCTGT	5791
Db	5401	TGTGACTGTGATGAGATGGGAAAGCGGCGCACAGACCCCGGGTCTCGTTGGCTGTCTGT	5460
0Y	5792	GGGCGCGCCCTGGCCACCTGTGTCTCTGGCTGTGAGGCTGCAGAGAGGCTCGTTCTGTGG	5851
Db	5461	GGGCGCGCCCTGGCCACCTGTGTCTCTGGCTGTGAGGCTGCAGAGAGGCTCGTTCTGTGG	5520
0Y	5852	TGGCGCGCTTCGCTCGGTTGGGCTGTGCTTACATTAACCGGTCGCAAGGCTCTGCA	5911
Db	5521	TGGCGCGCTTCGCTCGGTTGGGCTGTGCTTACATTAACCGGTCGCAAGGCTCTGCA	5580
0Y	5912	GGCCACTGTGAGACCTGGCTGCCCTGGGCAAGTGTCTCCGTTGTGAGACTGTGCTCAGGCC	5971
Db	5581	GGCCACTGTGAGACCTGGCTGCCCTGGGCAAGTGTCTCCGTTGTGAGACTGTGCTCAGGCC	5640
0Y	5972	AGGGCTCACAGCTGGGGCTCTGTCCGGAAGATGGGATCTTCTGGAGCTGCGCGGGA	6031
Db	5641	AGGGCTCACAGCTGGGGCTCTGTCCGGAAGATGGGATCTTCTGGAGCTGCGCGGGA	5700
0Y	6032	CAGAGTGGGGAGCTCTTAAGTTTGTGGGGGGAACCTTTGATATTCATGCGACGCTCCATCA	6091
Db	5701	CAGAGTGGGGAGCTCTTAAGTTTGTGGGGGGAACCTTTGATATTCATGCGACGCTCCATCA	5760
0Y	6092	CCCCACCCCTTTTGTGTACGACGACATAGTGTCTTACATTGGATTTTTGTAAAAAATATAA	6151
Db	5761	CCCCACCCCTTTTGTGTACGACGACATAGTGTCTTACATTGGATTTTTGTAAAAAATATAA	5820
0Y	6152	ATAAATGTAGAGACTTTTAAGTCT 6172	

```

|||||
DB      5821 AATTAATGAGACTTACTC 5841

RESULT 6
BC006363
LOCUS   Homo sapiens, exostoses (multiple)-like 3, clone MGC:12750
DEFINITION IMAGE:4131101, mRNA, complete cds.
ACCESSION BC006363
VERSION   BC006363.1 GI:13623512
KEYWORDS  MGC.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3241)
REFERENCE
AUTHORS  Strausberg, R.
TITLE     Direct Submission
JOURNAL   Submitted (09-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland.
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgcenhgti.nih.gov
Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McLoskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantipod, S., Thomas, P.J.,
Tiongson, E.E., Touchman, D.W., Tsurgoun, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 17 Row: 1 Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4103883.
Location/Qualifiers
1. 3241
/organism="Homo sapiens"
/db_xref="locustid:2137"
/db_xref="taxon:9606"
/clone="MGC:12750 IMAGE:4131101"
/tissue_type="Brain, neuroblastoma"
/clone_lib="NIH-MGC_19"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
122..2881
/codon_start=1
/product="exostoses (multiple)-like 3"
/protein_id="AAH06363.1"
/db_xref="GI:13623513"
/translation="MTGYMLKRGCGAGCGGCTGMLRNSRIRLITWSTFLVILVFFP
LTAHYLTITLDEADGKRIFGRVGNELCEVKAHVIDLCRIRSVSEELIOLEAKOE
LNSLEIAKLKLEACKRSIENAKODLQENKISQTESHYKEIMANOQKLSPIRL
PERKDAGLPKPRATRCRLHNCFDYSRCPITGFPYVYVSDQFVGSYLDPLVKOAF
QATARAANYVTENADIALCYLILVGEOMOPVYLRAPELEKOLYSLPHESTDGHNHYIT
NLSKSDNONLILYVSTGRAMAOSTFTYVOVRPGDLYVSPVHAHMSPPNMEIPPO
VYVKRYLFTPGGKTESIRSLQEARSEFEEMGGDPADYDRIITATLKAAYDSKLD
OVLVEPTCKNQKPSLPTEWALCGERDELLEKLTSTFADLITPGDPRLVISGCATR
LEALEEAVPVLVEQVOLPYQDMLOMNEALVYKPRVTEVHEFLRLSDSLIAM
RROGFMEVETVSTADISINTVIAMIRTIQIPAPIRREAAAEIPIHRSKAGADPN
MADNGDLIDGIVETEPYVAPRYLNFILITVTDYFRSMNCAAGCPHLPHPTFPDPLP

```

```

SEANFLSGTGRPDIGGAGGSGKEFQAALGNVPRQFTVVLTYEREVIANSLER
LNGDLYLNKVVVWNSPKLPSEDLIMPDIQVPIVWVREKSNLRFLPNEIEBAI
LISDDDAIRHDEIMFGRVREARDRIVGPRYHAMDIPOSMIXNSNSCETLSWY
LTGAFFHKYAYLYXYMROAIRDMNDEYNCDIAMEFNVSHTTRPPLKTVSRMT
FRCPGCPALSHDSSHFRERHKCNFVKYIGVIMPLLTORVDSVLFKTRLPDKTK
CFKFI"

BASE COUNT      702 a      925 c      898 g      716 t
ORIGIN

Query Match      52.1%; Score 3217; DB 9; Length 3241;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3220; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY  482 CACTACTCTTCTGGAACGCTGCTGTAACAGAGATCGTTTGTGAAATGCAACCA 541
DB  10 CACTACTCTTCTGGAACGCTGCTGTAACAGAGATCGTTTGTGAAATGCAACCA 69
QY  542 TGGTTATGCGAGTGACCCGAGCTGATCTGGGGGCGAGGCTGCAGAGACTCATGACAG 601
DB  70 TGGTTATGCGAGTGACCCGAGCTGATCTGGGGGCGAGGCTGCAGAGACTCATGACAG 129
QY  602 CTATACCATGCTGGGGGATGGGGGCGGCGGAGCAGGAGCTCATGCTGCGCTG 661
DB  130 CTATACCATGCTGGGGGATGGGGGCGGCGGAGCAGGAGCTCATGCTGCGCTG 189
QY  662 GTCCAAACCGCATCCGCGCTCAGCTGAGCTTCAGCTGCTTGTGATCCTGGCTTCTT 721
DB  190 GTCCAAACCGCATCCGCGCTCAGCTGAGCTTCAGCTGCTTGTGATCCTGGCTTCTT 249
QY  722 CCCGCTCATCGCCCACTATTACCTGACCATCTGTGATGAGGCTGATGAGGCAAGCG 781
DB  250 CCCGCTCATCGCCCACTATTACCTGACCATCTGTGATGAGGCTGATGAGGCAAGCG 309
QY  782 GATTTTGTGCGCGGGGGGGAACGAGCTGTGGAGGTGAAGACGCTGATCTGTG 841
DB  310 GATTTTGTGCGCGGGGGGGAACGAGCTGTGGAGGTGAAGACGCTGATCTGTG 369
QY  842 CCGCATCCGAGAGTGGGTGAGTGAAGAGCTCTGACCTGGAGGCCCAAGCGCAAGAGCT 901
DB  370 CCGCATCCGAGAGTGGGTGAGTGAAGAGCTCTGACCTGGAGGCCCAAGCGCAAGAGCT 429
QY  902 GAACAGGAGATGCCCAAGCTGAATCTGAAGATCGAAGCTGTGAAGAGCAATTGAGAA 961
DB  430 GAACAGGAGATGCCCAAGCTGAATCTGAAGATCGAAGCTGTGAAGAGCAATTGAGAA 489
QY  962 CGCCACAGAGACTGCTCCAGCTCAAGATGTATAGCCAGCAGCAGATTCCTCAAA 1021
DB  490 CGCCACAGAGACTGCTCCAGCTCAAGATGTATAGCCAGCAGCAGATTCCTCAAA 549
QY  1022 GGAGCTCATGAGCCAGAACAGCCCAAGCTGTCCTGCCATCGACTGCTCCAGAGAA 1081
DB  550 GGAGCTCATGAGCCAGAACAGCCCAAGCTGTCCTGCCATCGACTGCTCCAGAGAA 609
QY  1082 GGAGCATGCGGCGCTCCTCCCGGAAGGCACTCGGGGCTGCCGGCTACAACTGCTT 1141
DB  610 GGAGCATGCGGCGCTCCTCCCGGAAGGCACTCGGGGCTGCCGGCTACAACTGCTT 669
QY  1142 TGATTATTCGTTGGCCCTGCACTGCTGCTCCGCGCTGCTGCTGCTGCTGCTGCTGCTG 1201
DB  670 TGATTATTCGTTGGCCCTGCACTGCTGCTCCGCGCTGCTGCTGCTGCTGCTGCTGCTG 729
QY  1202 GTTGTGCTTGGAGAGTACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1261
DB  730 GTTGTGCTTGGAGAGTACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 789
QY  1262 AGCTAACGTTTATGTTACAGAAATGACAGATCGCTGCTTACGTGATCTAGTGGG 1321
DB  790 AGCTAACGTTTATGTTACAGAAATGACAGATCGCTGCTTACGTGATCTAGTGGG 849
QY  1322 AGAGATGACGAGGAGCGGCTGCTGCGGCTGCTGAGCTGAGAACAGATGATTCCTT 1381
DB  850 AGAGATGACGAGGAGCGGCTGCTGCGGCTGCTGAGCTGAGAACAGATGATTCCTT 909

```

QY 1382 GCCACTGGGGAGCGATGAGACAAACATGTCATCATCTGTCAGTAGTCAGA 1441
 |||||
 Db 910 GCCACTGGGGAGCGATGAGACAAACATGTCATCATCTGTCAGTAGTCAGA 969
 QY 1442 TACACAGACCTTCTCTATTAACGTCACTACTGCGCTGCCATGGTGGCCCACTCCACCTT 1501
 |||||
 Db 970 TACACAGACCTTCTCTATTAACGTCACTACTGCGCTGCCATGGTGGCCCACTCCACCTT 1029
 QY 1502 CTACACTGTCAGTACAGACCTGCTTGAAGTGTGATACAGCGGTGGTGCATGGCAT 1561
 |||||
 Db 1030 CTACACTGTCAGTACAGACCTGCTTGAAGTGTGATACAGCGGTGGTGCATGGCAT 1089
 QY 1562 GTCTGAGCCCCAACTTCAATGGAATCCACACAGAGTCCGGTGAAGCGGAATATCTTT 1621
 |||||
 Db 1090 GTCTGAGCCCCAACTTCAATGGAATCCACACAGAGTCCGGTGAAGCGGAATATCTTT 1149
 QY 1622 CACCTTCCAGGGCGAAGAAATGAGTCTCTAGAGTCTAGCTTACAGAGGCCCTCTCT 1681
 |||||
 Db 1150 CACCTTCCAGGGCGAAGAAATGAGTCTCTAGAGTCTAGCTTACAGAGGCCCTCTCT 1209
 QY 1682 CGAAGAGAAATGAGAGGGCGACCTCCCGCGACCTAGATACCGGATCATTTGCCACCT 1741
 |||||
 Db 1210 CGAAGAGAAATGAGAGGGCGACCTCCCGCGACCTAGATACCGGATCATTTGCCACCT 1289
 QY 1742 GAAGCGGTGACAGACAGCAAGTGTGATCAGGTCTGGTGAATTCACCTGCAAAAACA 1801
 |||||
 Db 1270 GAAGCGGTGACAGACAGCAAGTGTGATCAGGTCTGGTGAATTCACCTGCAAAAACA 1329
 QY 1802 GCCCAAAACCCAGCTGCCGACTGAGTGGGCACTGTGTGAGAGCGGAGACCCCTTGA 1861
 |||||
 Db 1330 GCCCAAAACCCAGCTGCCGACTGAGTGGGCACTGTGTGAGAGCGGAGACCCCTTGA 1389
 QY 1862 ATTGCTGAAGCTTCCACCTTCCGCTCATATTAACCCCGGGGACCTGCTGGTAT 1921
 |||||
 Db 1390 ATTGCTGAAGCTTCCACCTTCCGCTCATATTAACCCCGGGGACCTGCTGGTAT 1449
 QY 1922 TTCTCTGGGTGTGCAACACAGGCTTTCGAAGCCCTGGAATCGGTGCCGTGCCGTGGT 1981
 |||||
 Db 1450 TTCTCTGGGTGTGCAACACAGGCTTTCGAAGCCCTGGAATCGGTGCCGTGCCGTGGT 1509
 QY 1982 GCTGGGGAGACAGTCCAGCTTCCCTACAGAGACATGCTGAGTGAAGCAGGCGGCCCT 2041
 |||||
 Db 1510 GCTGGGGAGACAGTCCAGCTTCCCTACAGAGACATGCTGAGTGAAGCAGGCGGCCCT 1569
 QY 2042 GGTGGTGCACAAAGCCTGCTGTACAGAGTTCATTTCTGCTCAGAAACCTCTCGATAG 2101
 |||||
 Db 1570 GGTGGTGCACAAAGCCTGCTGTACAGAGTTCATTTCTGCTCAGAAACCTCTCGATAG 1629
 QY 2102 TGACCTCTGCTATGAGGCGGCAAGGCGCTTCTCTGGGAGACTTACTTCTCCACTGC 2161
 |||||
 Db 1630 TGACCTCTGCTATGAGGCGGCAAGGCGCTTCTCTGGGAGACTTACTTCTCCACTGC 1689
 QY 2162 TGACAGTATTTTATTAATACCGTGTGCTGCTATGATTAAGACTCGCATCCAGCCGCC 2221
 |||||
 Db 1690 TGACAGTATTTTATTAATACCGTGTGCTGCTATGATTAAGACTCGCATCCAGCCGCC 1749
 QY 2222 TCCCATCCGGGAGAGGCGGAGCTGAGATCCCCACCGTTCAGGCAAGGCGGCTGGAAC 2281
 |||||
 Db 1750 TCCCATCCGGGAGAGGCGGAGCTGAGATCCCCACCGTTCAGGCAAGGCGGCTGGAAC 1809
 QY 2282 TGACCCCAACATGAGCTGACAAAGGGGAGCTGAGCTGGGGCCAGTGAAGCAGGAGCCGC 2341
 |||||
 Db 1810 TGACCCCAACATGAGCTGACAAAGGGGAGCTGAGCTGGGGCCAGTGAAGCAGGAGCCGC 1869
 QY 2342 CTAGCGCTCACCAGATACCTCCGCAATTTCACTGTACTGTCACTGCTTATTAACCGGAG 2401
 |||||
 Db 1870 CTAGCGCTCACCAGATACCTCCGCAATTTCACTGTACTGTCACTGCTTATTAACCGGAG 1929
 QY 2402 CTGGAAGTGTCTCAGGCGCTTTCACATCTTTCGCCACACTCCCTTGAACCTGTGT 2461
 |||||
 Db 1930 CTGGAAGTGTCTCAGGCGCTTTCACATCTTTCGCCACACTCCCTTGAACCTGTGT 1989

QY 2462 GCCCTCAGAGCCCAATTTCTTGGGCTAGGAGCTGCTTTCCGGCTATTTGTGTGGAGC 2521
 |||||
 Db 1990 GCCCTCAGAGCCCAATTTCTTGGGCTAGGAGCTGCTTTCCGGCTATTTGTGTGGAGC 2049
 QY 2522 TGGGGGTTTGGCAAGGAATTTCCAGGAGGCTTGGAGGCAATTTGCCGAGAGCAGTT 2581
 |||||
 Db 2050 TGGGGGTTTGGCAAGGAATTTCCAGGAGGCTTGGAGGCAATTTGCCGAGAGCAGTT 2109
 QY 2582 CACGGTGTGATGTTGACTTATGAGCGGAGAGAGTCTTATGAACCTTTTAGAGAGCT 2641
 |||||
 Db 2110 CAGGCTGTGATGTTGACTTATGAGCGGAGAGAGTCTTATGAACCTTTTAGAGAGCT 2169
 QY 2642 GAATGCGCTCCCTTACCTGAAACAGTGTGTGTGTGTGGAATTTCCCAAGCTGCCATC 2701
 |||||
 Db 2170 GAATGCGCTCCCTTACCTGAAACAGTGTGTGTGTGTGGAATTTCCCAAGCTGCCATC 2229
 QY 2702 AGAGAGCTTCTGTGGGCTGACATGTGGCTGTCCATCATGATGATGATGCTGTGGAAGAA 2761
 |||||
 Db 2230 AGAGAGCTTCTGTGGGCTGACATGTGGCTGTCCATCATGATGATGATGCTGTGGAAGAA 2289
 QY 2762 CAGTTTGAACACCGATTTCTTACCTGGAATGAATTTGAGACAGAGGCCATCTGTCCAT 2821
 |||||
 Db 2290 CAGTTTGAACACCGATTTCTTACCTGGAATGAATTTGAGACAGAGGCCATCTGTCCAT 2349
 QY 2822 TGATGACGATGCTACCTCCGCCATGACGAAATATGTTTGGGTTCCGGGTGTGAGAGA 2881
 |||||
 Db 2350 TGATGACGATGCTACCTCCGCCATGACGAAATATGTTTGGGTTCCGGGTGTGAGAGA 2409
 QY 2882 AGCTCGGGAGCGCATGCTGGGGCTTCCCGGCGTTTACACAGCATGGGACATCCCCCATCA 2941
 |||||
 Db 2410 AGCTCGGGAGCGCATGCTGGGGCTTCCCGGCGTTTACACAGCATGGGACATCCCCCATCA 2469
 QY 2942 GTCCGTGCTTACAACTCAACTCAACTCTCTGTGAGCTGTCTGATGAGTGTGTC 3001
 |||||
 Db 2470 GTCCGTGCTTACAACTCAACTCAACTCTCTGTGAGCTGTCTGATGAGTGTGTC 2529
 QY 3002 CTCTTTTACAAATTAATATGCTTACCTGTATTTATGTATGTCGCCAGGCCATCCGGA 3061
 |||||
 Db 2530 CTCTTTTACAAATTAATATGCTTACCTGTATTTATGTATGTCGCCAGGCCATCCGGA 2589
 QY 3062 CATGTGGATTAATCAATCAACTGTGAGAGACATGGCATGAACCTCTGTCTCCCAT 3121
 |||||
 Db 2590 CATGTGGATTAATCAATCAACTGTGAGAGACATGGCATGAACCTCTGTCTCCCAT 2649
 QY 3122 CACTCGGAAGCCCCCAATCAAGGTGACCTCAAGGTGACATTTCCGATGCCCCAGAGTGC 3181
 |||||
 Db 2650 CACTCGGAAGCCCCCAATCAAGGTGACCTCAAGGTGACATTTCCGATGCCCCAGAGTGC 2709
 QY 3182 TCAGGCCCTGTCTATGATGACTCCCACTTCCACGAGCGCACAAATGCTCAACTCTT 3241
 |||||
 Db 2710 TCAGGCCCTGTCTATGATGACTCCCACTTCCACGAGCGCACAAATGCTCAACTCTT 2769
 QY 3242 CGTGAAGTGTAGGGCTAATCAATGCCCCCTCTGTACAGGAGTTCAAGGGGATTTGTGCT 3301
 |||||
 Db 2770 CGTGAAGTGTAGGGCTAATCAATGCCCCCTCTGTGTACAGGAGTTCAAGGGGATTTGTGCT 2829
 QY 3302 CTTCAAGACAGCCCTGCCCCATGACAGAACCAAGTGTCTCAAGTTCTTAGGGGCAAGC 3361
 |||||
 Db 2830 CTTCAAGACAGCCCTGCCCCATGACAGAACCAAGTGTCTCAAGTTCTTAGGGGCAAGC 2889
 QY 3362 CACGGTCTGGGGAAGAGATGAGCAGAGGAGGAAGTGTCTCCCAAGTTCCTTAGGCAT 3421
 |||||
 Db 2890 CACGGTCTGGGGAAGAGATGAGCAGAGGAGGAAGTGTCTCCCAAGTTCCTTAGGCAT 2949
 QY 3422 TGCAAGACCTTGGGCAATCTGCTGTGGGGGCGGCAAGGCTGCTGGAAGGGGAGC 3481
 |||||
 Db 2950 TGCAAGACCTTGGGCAATCTGCTGTGGGGGCGGCAAGGCTGCTGGAAGGGGAGC 3009
 QY 3482 AGGAGAGTGTGAAGAAACCGCTCTTATCTTGAAGTCAAGGCACACTGGGCTGTGAGC 3541
 |||||
 Db 3010 AGGAGAGTGTGAAGAAACCGCTCTTATCTTGAAGTCAAGGCACACTGGGCTGTGAGC 3069
 QY 3542 CTTGGGCGGAGTCCCCCGGGTTCCCAACACAGGGCACTGACTGATACCTTACACTGAGGA 3601

|||||
Db 3070 CTTGGGGAGTCCCGGGGTTCCCAACAGGGGCTGACTGATTAAGCTTAACACTGAGGA 3129
|||||
QY 3602 CTGTGGGACTCTGCAGAGTCACTACACCGTTGTACGGCCAGGAGAGCTGTCTGG 3661
|||||
Db 3130 CTGTGGGACTCTGCAGAGTCACTACACCGTTGTACGGCCAGGAGAGCTGTCTGG 3189
|||||
QY 3662 TTTTTCATTCAATTAACAACTATTATGATTATTTAAAGAGAA 3706
|||||
Db 3190 TTTTTCATTCAATTAACAACTATTATGATTATTTAAAGAGAA 3234
|||||
RESULT 7
AX332231/c 167343 bp DNA linear PAT 09-JAN-2002
LOCUS AX332231
DEFINITION Sequence 2740 from Patent WO0194629.
ACCESSION AX332231
VERSION AX332231.1 GI:18122865
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young P.E., Augustus M., Carter K.C., Edner R., Endress G.,
Horrigian S., Soppet D.R. and Weaver Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 2740 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source 1. 167343
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 49550 a 37098 c 35753 g 44929 t 13 others
ORIGIN
Query Match 49.2%; Score 3035; DB 6; Length 167343;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 3062; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 3066 GTGATGATATACATCACTGTGAGACATTCGCACTCTCTTCTCCACATCACT 3125
|||||
Db 67898 GTGTGGGATCGGCTCAGCTGCAAGGCTTGTCTCACTATTACCTGGATTCACCTGTGTC 67839
|||||
QY 3126 CGGAAGCCCCCATCAAGTGAAGTCACTCAAGTGAAGTCACTCACTCACTCACT 3185
|||||
Db 67838 TCTCTCCGCTTCCAGGTGACCTCAAGGTGACATTCGATGCCAGGATGCCCTCAG 67779
|||||
QY 3186 GCCCTGTCTATGATGACTCCACTTCCAGAGGGGACAAAGTCACTCACTCTGCTG 3245
|||||
Db 67778 GCCCTGTCTATGATGACTCCACTTCCAGAGGGGACAAAGTCACTCACTCTGCTG 67719
|||||
QY 3246 AAGGTGATGAGCTATCACTCCCTCTCTCACTCACTCACTCACTCACTCACTCT 3305
|||||
Db 67718 AAGGTGATGAGCTATCACTCCCTCTCTCACTCACTCACTCACTCACTCACTCT 67659
|||||
QY 3306 AAGACACGCTGCCCCATGACAAAGACCAAGTGTCAAGTCACTCACTCACTCACT 3365
|||||
Db 67658 AAGACACGCTGCCCCATGACAAAGACCAAGTGTCAAGTCACTCACTCACTCACT 67599
|||||
QY 3366 GTCTGGGGAAG 3425
|||||
Db 67598 GTCTGGGGAAG 67539
|||||
QY 3426 GGAACCTTGGGCAATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3485
|||||
Db 67538 GGAACCTTGGGCAATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 67479
|||||
QY 3486 GGAAGTGAAGGAAG 3545
|||||
Db 67478 GGAAGTGAAGGAAG 67419

QY 3546 GGGGAGATCCCCGGGGTTCCCAACAGGGGCTGACTGATTAAGCTTAACACTGAGAGACTGT 3605
|||||
Db 67418 GGGGAGATCCCCGGGGTTCCCAACAGGGGCTGACTGATTAAGCTTAACACTGAGAGACTGT 67359
|||||
QY 3606 GGGGACTCTGCAGAGTCACTACACCGTTGTACGGCCAGGAGAGAGCTGTCTGGTTT 3665
|||||
Db 67358 GGGGACTCTGCAGAGTCACTACACCGTTGTACGGCCAGGAGAGAGCTGTCTGGTTT 67299
|||||
QY 3666 TACATTCATTAACAACTATTATGATTATTTAAAGAGAAAGTTGATGATTGCCATTC 3725
|||||
Db 67298 TACATTCATTAACAACTATTATGATTATTTAAAGAGAAAGTTGATGATTGCCATTC 67239
|||||
QY 3726 AGGCTATTAT 3785
|||||
Db 67238 AGGCTATTAT 67179
|||||
QY 3786 ATTTTGGCTGGGGAGTGGAGTTTGGCTTCTTAAGGAGAGAGAGAGAGAGAGAGAGCT 3845
|||||
Db 67178 ATTTTGGCTGGGGAGTGGAGTTTGGCTTCTTAAGGAGAGAGAGAGAGAGAGAGAGCT 67119
|||||
QY 3846 TGTCTGATTTCTGGGAGATGGGTCTGGCTTGTCTCACTGGCTTATCTTAAGAT 3905
|||||
Db 67118 TGTCTGATTTCTGGGAGATGGGTCTGGCTTGTCTCACTGGCTTATCTTAAGAT 67059
|||||
QY 3906 CATTCCCATCTCTCCCAAGCCCATCTGTGTGACAGCAACAGAAAGGATGAACTTGGCC 3965
|||||
Db 67058 CATTCCCATCTCTCCCAAGCCCATCTGTGTGACAGCAACAGAAAGGATGAACTTGGCC 66999
|||||
QY 3966 CTCTTGGGGGCTGGACAAGTCTCTCTTCACTTCACTTCTGTGCAAGTCAAGCACTGT 4025
|||||
Db 66998 CTCTTGGGGGCTGGACAAGTCTCTCTTCACTTCACTTCTGTGCAAGTCAAGCACTGT 66939
|||||
QY 4026 AACTCAATCTCTTCCCAAGTGAATCCCTGGAGGCGCTGACCCCTGTGTGGCTTCTCAGC 4085
|||||
Db 66938 AACTCAATCTCTTCCCAAGTGAATCCCTGGAGGCGCTGACCCCTGTGTGGCTTCTCAGC 66879
|||||
QY 4086 TTCCCTGCTGGGGGCGGAGGATTTTGAAGATTTATCTTTAGGCGAGGCTTGGCTCCGT 4145
|||||
Db 66878 TTCCCTGCTGGGGGCGGAGGATTTTGAAGATTTATCTTTAGGCGAGGCTTGGCTCCGT 66819
|||||
QY 4146 AACTATCCCTCTCTCCCAATTTCTCTTGTGAGAGATGAGAGAGAGAGAGAGAGAG 4205
|||||
Db 66818 AACTATCCCTCTCTCCCAATTTCTCTTGTGAGAGATGAGAGAGAGAGAGAGAGAG 66759
|||||
QY 4206 GAAAGATAGGGGCTGAAGAGCCCACTCCAGATGGCTCTTCTATCTGCTCTCTGTT 4265
|||||
Db 66758 GAAAGATAGGGGCTGAAGAGCCCACTCCAGATGGCTCTTCTATCTGCTCTCTGTT 66699
|||||
QY 4266 GAAACACACGCTGTGGGCTCAAGGCTTCTCAAGTGTCTCTTCTGATTTGGACAGG 4325
|||||
Db 66698 GAAACACACGCTGTGGGCTCAAGGCTTCTCAAGTGTCTCTTCTGATTTGGACAGG 66639
|||||
QY 4326 AGATCAGCAGCTGACATCTGCTGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGT 4385
|||||
Db 66638 AGATCAGCAGCTGACATCTGCTGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGT 66579
|||||
QY 4386 CTAGTGAAGCAAGCAGTGTCTTTCAGAGAAACCAAGGCTGAGGAGAGAGTGTTC 4445
|||||
Db 66578 CTAGTGAAGCAAGCAGTGTCTTTCAGAGAAACCAAGGCTGAGGAGAGAGTGTTC 66519
|||||
QY 4446 AGCAAGGCGCTGCTTGGATTAATCTGTGTGAATTCACCTTCCCGGCTCTGTGTC 4505
|||||
Db 66518 AGCAAGGCGCTGCTTGGATTAATCTGTGTGAATTCACCTTCCCGGCTCTGTGTC 66459
|||||
QY 4506 TGGAGCCCATCTCTGTATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4565
|||||
Db 66458 TGGAGCCCATCTCTGTATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 66399
|||||
QY 4566 ACTCCCGAGGTTTGTATGTCTAGAACAAATGGAGGCTGTGATTTGGTGTAGCTTC 4625
|||||
Db 66398 ACTCCCGAGGTTTGTATGTCTAGAACAAATGGAGGCTGTGATTTGGTGTAGCTTC 66339

4626 ACATCCAGCCTTGGAACTCTAACGGGCAATTCACAACCCGAGTTACCACTTTCCACTCCCTG 4685
|||||
Db 66338 ACATCCAGCCTTGGAACTCTAACGGGCAATTCACAACCCGAGTTACCACTTTCCACTCCCTG 66379
4686 CTTAGGATTCGTCCCTGGGCTGAACCTGAATAGCTAATTTTGGGTCAGGTCG 4745
|||||
Db 66278 CTTAGGATTCGTCCCTGGGCTGAACCTGAATAGCTAATTTTGGGTCAGGTCG 66219
4746 AGTAGGGGAACCTAGAGAGGTGTAGTGGCAATTTGTCAAGGATTTAGCCATGACGTGT 4805
|||||
Db 66218 AGTAGGGGAACCTAGAGAGGTGTAGTGGCAATTTGTCAAGGATTTAGCCATGACGTGT 66159
4806 TCTTGAACCCCTACTTTCTGGGAAGTGGAGTTACTCTGGAATTTTCTAGCAACTGAAACA 4885
|||||
Db 66158 TCTTGAACCCCTACTTTCTGGGAAGTGGAGTTACTCTGGAATTTTCTAGCAACTGAAACA 66099
4866 AAGCTCAGGTTTGTCTGCTGATGACATGCTTAAAGCAGTTCCGCTTCCCTAGACCT 4925
|||||
Db 66098 AAGCTCAGGTTTGTCTGCTGATGACATGCTTAAAGCAGTTCCGCTTCCCTAGACCT 66039
4926 TGGCATCTGTGCTTCTAATTTCTTGAATATAGTTCTCCTGTGACCTGCTGTACACGTG 4985
|||||
Db 66038 TGGCATCTGTGCTTCTAATTTCTTGAATATAGTTCTCCTGTGACCTGCTGTACACGTG 65979
4986 GGTCTCTTCAAGTACTGTTTGAAGCTGGGCTTTTGTGTAGCTCCACACCTGTA 5045
|||||
Db 65978 GGTCTCTTCAAGTACTGTTTGAAGCTGGGCTTTTGTGTAGCTCCACACCTGTA 65919
5046 GGGGTAGTGGGCTTAAAGGAACCTCCCATTTGGCAACGGGACGGGCGGCGCCAG 5105
|||||
Db 65918 GGGGTAGTGGGCTTAAAGGAACCTCCCATTTGGCAACGGGACGGGCGGCGCCAG 65859
5106 ACTGTGTTTCCAAAGGTTCGCCGCCGCCAACCCAGCATCAGCTGTAGTCCCTGCTG 5165
|||||
Db 65858 ACTGTGTTTCCAAAGGTTCGCCGCCGCCAACCCAGCATCAGCTGTAGTCCCTGCTG 65799
5166 AGGCAAGTGGTATATGTTTCCAGAGTGGGGGTGAGAGCCCTTCTCAGAACTTTCTAG 5225
|||||
Db 65798 AGGCAAGTGGTATATGTTTCCAGAGTGGGGGTGAGAGCCCTTCTCAGAACTTTCTAG 65739
5226 TTGGCTCTTACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5285
|||||
Db 65738 TTGGCTCTTACCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 65679
5286 AGCCAAAGAGTGTGTGTGGGCTATATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5345
|||||
Db 65678 AGCCAAAGAGTGTGTGTGGGCTATATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 65619
5346 ATGTGAGGAATCTCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 5405
|||||
Db 65618 ATGTGAGGAATCTCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 65559
5406 TGGGACCCCTTTTAAAGGAGTCACTAGTGGGCTGTGTGTGGGAGGAGGAGGAGGAGGAG 5465
|||||
Db 65558 TGGGACCCCTTTTAAAGGAGTCACTAGTGGGCTGTGTGTGGGAGGAGGAGGAGGAGGAG 65499
5466 TCTCCCTTTTAAAGGAGTGTGTGGGCTGTGTGTGGGAGGAGGAGGAGGAGGAGGAGGAG 5525
|||||
Db 65498 TCTCCCTTTTAAAGGAGTGTGTGGGCTGTGTGTGGGAGGAGGAGGAGGAGGAGGAGGAG 65439
5526 CTGAGTGTGTGACACCAAAACATGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5585
|||||
Db 65438 CTGAGTGTGTGACACCAAAACATGCGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 65379
5586 AGGACAGCCCTCTTGAACCTGGAAGGAACACTGGCTTGAAGGAGTGGAGAGAGGAGGAGGAG 5645
|||||
Db 65378 AGGACAGCCCTCTTGAACCTGGAAGGAACACTGGCTTGAAGGAGTGGAGAGAGGAGGAGGAG 65319
5646 AGGAGGAGCCCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5705
|||||
Db 65318 AGGAGGAGCCCTCTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 65259
5706 TTTCACACGAGGGAAGGCGCGGAGCTGTGACTGCTTGAATGGGAAGGCGCGCCACA 5765

|||||
Db 65258 TTTCACACGAGGGAAGGCGCGGAGCTGTGACTGCTTGAATGGGAAGGCGCGCCACA 65199
5766 GACCCGGGCTCTCTTGGCTGTGTGTGGGCGCCCTGTGGACCTTGTCTGTGGCTGAG 5825
|||||
Db 65198 GACCCGGGCTCTCTTGGCTGTGTGTGGGCGCCCTGTGGACCTTGTCTGTGGCTGAG 65139
5826 GTGTCAGAGGCGCTCTCTGTGTGTGGGCGCGGCTGTGTGTGGGCTGTGTGTGTGTGTGT 5885
|||||
Db 65138 GTGTCAGAGGCGCTCTCTGTGTGTGGGCGCGGCTGTGTGTGGGCTGTGTGTGTGTGTGT 65079
5886 CCATTAACCGTCCAGGCTGTGAGGCACTGTGAGCGCTGTGCTGTGCTGTGCTGTGCTGTGCT 5945
|||||
Db 65078 CCATTAACCGTCCAGGCTGTGAGGCACTGTGAGCGCTGTGCTGTGCTGTGCTGTGCTGT 65019
5946 CCTCCGTGTGAGCTGTGCTGTGAGGCGCAGGCTGTGAGGCTGTGCTGTGCTGTGCTGTGCT 6005
|||||
Db 65018 CCTCCGTGTGAGCTGTGCTGTGAGGCGCAGGCTGTGAGGCTGTGCTGTGCTGTGCTGTGCT 64959
6006 GGGATCTTCTGTGGAGCTGCGCGGAGAGAGTGGGAGCTCTGATTGTGGGGGAGAGC 6065
|||||
Db 64958 GGGATCTTCTGTGGAGCTGCGCGGAGAGAGTGGGAGCTCTGATTGTGGGGGAGAGC 64899
6066 TTGTATATCCATGCGCAGCTCATCCACCCCTTTTGTGTACGAGCACAATGTCTT 6125
|||||
Db 64898 TTGTATATCCATGCGCAGCTCATCCACCCCTTTTGTGTACGAGCACAATGTCTT 64839
6126 ACATTGGATTTTGTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 6172
|||||
Db 64838 ACATTGGATTTTGTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 64792

RESULT 8
AX335067/c AX335067 167343 bp DNA linear PAT 09-JAN-2002
LOCUS Sequence 5576 from Patent WO0194629.
DEFINITION AX335067
ACCESSION AX335067
VERSION AX335067.1 GI:18125786
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horligan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 5576 13-DEC-2001;
FEATURES
source 1..167343
Location/Qualifiers
1..167343
"organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 49550 a 37098 c 35753 g 44929 t 13 others
ORIGIN

Query Match 49.2%; Score 3035; DB 6; Length 167343;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 3062; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 3066 GTGATGAAATCAATCACTGTGAGACATTCGATGAATCTTGTCTGCCACATCACT 3125
|||||
Db 67898 GTGATGAAATCAATCACTGTGAGACATTCGATGAATCTTGTCTGCCACATCACT 67839
Qy 3126 CGGAAGCCCTCAATCAAGGTACTACAGGTGTGAGCATTTCCAGATGCCAGATG 3185
|||||
Db 67838 TCTCTCTCCCTTTCAGGTACTACAGGTGTGAGCATTTCCAGATGCCAGATG 67779
Qy 3186 GCCCTGTCTATGATGATCCCACTTCCAGAGGAGGAGCAAGTGCATCAATCTTCTG 3245
|||||
Db 67778 GCCCTGTCTATGATGATCCCACTTCCAGAGGAGGAGCAAGTGCATCAATCTTCTG 67719


```
|||||
Db 65558 TGGGACCCCTTTAGCGGAGTCAAGTAGCGGGGCCCTGTGTGGGAGCCAGGCC 65499
OY 5466 TCCTCCCTTTAGTGGAGCAGTTGTCGGGGCCCGGAATGTCACCTGGTGGATCTAAGAGGG 5525
Db 65498 TCTCCCTTTAGTGGAGCAGTTGTCGGGGCCCGGAATGTCACCTGGTGGATCTAAGAGGG 65439
OY 5526 CTGAGTGGTGTGACACCAAAACATGCGCAGAGGGGCTGTGTGCGCGCTTCCACA 5585
Db 65438 CTGAGTGGTGTGACACCAAAACATGCGCAGAGGGGCTGTGTGCGCGCTTCCACA 65379
OY 5586 AGGACAGCCCTCTTGTGACCTGGAAGAACACTGGCTTGAAGACTGCAGACAGGCTCTG 5645
Db 65378 AGGACAGCCCTCTTGTGACCTGGAAGAACACTGGCTTGAAGACTGCAGACAGGCTCTG 65319
OY 5646 AGGGGACAGCCCTCTGAGGAGAGGAGCAAGGTGGGACAGTCACTGTGAGGAGTGC 5705
Db 65318 AGGGGACAGCCCTCTGAGGAGAGGAGCAAGGTGGGACAGTCACTGTGAGGAGTGC 65359
OY 5706 TTCTCACACGAGGAAGCCCGCAGCTGTGACTGCTGTGAGATGGAAAGCGGCCACA 5765
Db 65258 TTCTCACACGAGGAAGCCCGCAGCTGTGACTGCTGTGAGATGGAAAGCGGCCACA 65199
OY 5766 GACCCCGGCTCTCTGGCTGTCTGTGGGCGCCCGCTGGCCTTGTCTGTGGCTCCGAG 5835
Db 65198 GACCCCGGCTCTCTGGCTGTCTGTGGGCGCCCGCTGGCCTTGTCTGTGGCTCCGAG 65139
OY 5826 GGTGACAGAGCCCTGCTGCTGTGGTGGCGGCTGCTGCTGGCTTGTGGCTGTCTTA 5885
Db 65138 GGTGACAGAGCCCTGCTGCTGTGGTGGCGGCTGCTGCTGGCTTGTGGCTGTCTTA 65079
OY 5886 CCATTAACACCTCCAGAGGCTCTGCAGAGCCACTGTGAGCGCTGCTCCCTGGCAGTGC 5945
Db 65078 CCATTAACACCTCCAGAGGCTCTGCAGAGCCACTGTGAGCGCTGCTCCCTGGCAGTGC 65019
OY 5946 CCTCGGTGTGACGTGCTGTGAGGCGAGGCTCAGCGCTGCTGCTGGTCCGGAAGAT 6005
Db 65018 CCTCGGTGTGACGTGCTGTGAGGCGAGGCTCAGCGCTGCTGCTGGTCCGGAAGAT 64959
OY 6006 GGGATCTTTCTGGAGCTGCGCCGACAGAGTGGGAGACTCTAGTTGTGGGGGAAGC 6065
Db 64958 GGGATCTTTCTGGAGCTGCGCCGACAGAGTGGGAGACTCTAGTTGTGGGGGAAGC 64899
OY 6066 TTTATATCATGTCAGCTCATCCACCCCTTTTGTCTCAGACACATAATGCTCTT 6125
Db 64898 TTTATATCATGTCAGCTCATCCACCCCTTTTGTCTCAGACACATAATGCTCTT 64839
6126 ACATTGATTTTGTAAAAATAATAATGAGACTTAATCTC 6172
64838 ACATTGATTTTGTAAAAATAATAATGAGACTTAATCTC 64792
|||||
RESULT 9
HSU96629/c 167343 bp DNA linear PRI 10-JAN-2000
LOCUS HSU96629
DEFINITION Human chromosome 8 BAC clone C1987SK-2A8 complete sequence.
ACCESSION U96629.1 GI:2341008
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 167343)
AUTHORS Loftus,B.J., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R.,
Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
Deslattes,Mays,A., Cao,X., Xu,R.X., Kang,H.L., Mitchell,S.,
Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.
Genome duplications and other features in 12 Mb of DNA sequence
from human chromosome 16p and 16q
JOURNAL Genomics 60 (3), 295-308 (1999)
MEDLINE 99425270
PUBMED 10493829
```

```
REFERENCE 2 (bases 1 to 167343)
AUTHORS Adams,M.D.
TITLE Human chromosome 8 BAC clone C1987SK-2A8 complete sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 167343)
AUTHORS Adams,M.D.
TITLE Direct Submission
SUBMITTED (07-APR-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
4 (bases 1 to 167343)
REFERENCE Adams,M.D., Loftus,B.J., Zhou,L., La Bombard,M., Kim,U.J. and
Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Aug 22, 1997 this sequence version replaced gi:1930148.
BAC clone C1987SK-2A8 is located on chromosome 8. Genes were
identified by a combination of five methods: XGRAIL (available by
anonymous ftp from arthur.epm, GeneFinder (available by anonymous
ftp from collins.washington.edu), GENSCAN (available e-mail server
at genscan@genomic.stanford.edu), searches of the EST database at
TIGR (http://www.tigr.org/tdb/hcd/hcd.html) and searches against a
peptide database. Repeats were identified using Censor (Jurke, J.,
Klonowski, P., Dagman, V., Pelton, P. Censor a program for the
identification and elimination of repetitive elements from DNA
sequences. Computers Chem 20: 119-121 (1996), available by anonymous
ftp from ncbl.nlm.nih.gov).
location/Qualifiers
1..167343
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/clone="2A8"
/complement(20..308)
/rpt_family="Alu-Jo"
complement(370..1051)
/rpt_family="LI"
1070..1355
/rpt_family="Alu-Jo"
complement(1382..2775)
/rpt_family="LI"
complement(3086..3283)
/rpt_family="Alu-Jo"
3861..3991
/rpt_family="Alu-Jo"
complement(4198..4490)
/rpt_family="Alu-Jb"
complement(4878..5157)
/rpt_family="Alu-S2"
complement(5650..5945)
/rpt_family="Alu-Sx"
complement(7305..7586)
/rpt_family="Alu-Sz"
complement(7613..7717)
/rpt_family="LIMB3"
8787..9080
/rpt_family="Alu-Jo"
complement(9234..9495)
/rpt_family="LIMB3"
complement(9630..9902)
/rpt_family="Alu-Y"
complement(9914..10017)
/rpt_family="LIMB7"
10304..10598
/rpt_family="Alu-Sq"
11339..11669
/rpt_family="Alu-Jb"
complement(12169..12457)
/rpt_family="Alu-Sz"
complement(12844..13048)
/rpt_family="MLT1B"
complement(13100..13188)
/rpt_family="Alu-J"
```

FEATURES

source

repeat_region	complement(13790..14082)
repeat_region	/rpt_family="Alu-Sx"
repeat_region	complement(14138..14415)
repeat_region	/rpt_family="Alu-Jo"
repeat_region	complement(14496..15257)
repeat_region	/rpt_family="Alu-Jo"
repeat_region	complement(15294..15582)
repeat_region	/rpt_family="Alu-Sx"
repeat_region	complement(15906..15984)
repeat_region	/rpt_family="MER31"
repeat_region	16308..16403
repeat_region	/rpt_family="MIR"
repeat_region	16420..16580
repeat_region	/rpt_family="L1"
repeat_region	16598..16694
repeat_region	/rpt_family="Alu-Jb"
repeat_region	16695..16975
repeat_region	/rpt_family="Alu-Jo"
repeat_region	complement(17050..17344)
repeat_region	/rpt_family="Alu-Sx"
repeat_region	complement(17582..17871)
repeat_region	/rpt_family="Alu-Sz"
repeat_region	18635..18924
repeat_region	/rpt_family="Alu-Jb"
repeat_region	complement(19052..19565)
repeat_region	/rpt_family="MLT2C2"
repeat_region	complement(19711..19999)
repeat_region	/rpt_family="Alu-Sx"
repeat_region	complement(20040..20356)
repeat_region	/rpt_family="Alu-Sx"
repeat_region	20613..20713
repeat_region	/rpt_family="Alu-Jo"
repeat_region	20851..21135
repeat_region	/rpt_family="Alu-Sz"
repeat_region	21254..21434
repeat_region	/rpt_family="MERSA"
repeat_region	21515..21858
repeat_region	/rpt_family="THE1B"
repeat_region	join(21859..21967,24521..24701,30613..30673,37496..37667,40528..40652,42553..42720,47399..47497,48453..48590,50157..50816)
repeat_region	/gene="2A8.2"
repeat_region	/gene="2A8.2"
repeat_region	21859..50816
repeat_region	/gene="2A8.2"
repeat_region	join(21859..21967,24521..24701,30613..30673,37496..37667,40528..40652,42553..42720,47399..47497,48453..48590,50157..50333)
repeat_region	/gene="2A8.2"
repeat_region	/codon_start=1
repeat_region	/product="unknown protein CTF987SK_2A8_1"
repeat_region	/protein_id="BAB67601.1"
repeat_region	/db_xref="GI:1930149"
repeat_region	/translation="MDQASLNSDVLVTGILQIPTANPDGMGEFCSNALITVRNGANLVPCPSGVGYDLDLQKIDYIDSAGLSVPLFIISPVANSLEFQIFAEMLCNKKOSKVALPEPPFPAHELQITNKLKHPYSIHGDFSNDFROPVCVFTFGHPSLNRGDVHFMELMGSLSNLWTVIFTEPDESYLEALAPYQPLAMKCTICPIDRLNFQVSKLKEVQPLHVHCEQOTQPPAQSHRMDLMDQPPAMSYKRAEVLALPFKKRYKTEIMPELASLTPMKIKGISATYSALVHTRKDKNKHLLQPPRPAQPTSGKKRRVSDVDVDCVKLPILSSGTFPQPTLEKHFSDIKVEDYDAKHHIVLQEAETLLIQIEDSTHIIQDNDEMKRVLRLVLFKEQK"
repeat_region	21967..22392
repeat_region	/rpt_family="MLT2B2"
repeat_region	complement(22810..23089)
repeat_region	/rpt_family="Alu-Jb"
repeat_region	24436..24747
repeat_region	/rpt_family="Alu-Sz"
repeat_region	complement(25510..25694)
repeat_region	/rpt_family="MERSA"
repeat_region	27021..27409
repeat_region	/rpt_family="MSTA"
repeat_region	27886..28172
repeat_region	/rpt_family="Alu-Sc"
repeat_region	28186..28457

		Query Match Best Local Similarity	Matches 3062;	Conservative	49.2%; 98.6%;	Score 3035; Pred. No. 0;	DB 9;	Length 167343;	
							Mismatches 45;	Indels	Gaps
	repeat_region	/rpt_family="Alu-Jo"	28831.	.29113					
	repeat_region	/rpt_family="Alu-Sx"	complement(29159..	.29615)					
	repeat_region	/rpt_family="LlMA9"	complement(29616..	.29892)					
	repeat_region	/rpt_family="Alu-Sg"	complement(29911..	.29941)					
	repeat_region	/rpt_family="LlMA9"	29965..	.30243					
	repeat_region	/rpt_family="Alu-Sx"	complement(30252..	.30477)					
	repeat_region	/rpt_family="LlMA10"	30492..	.30781					
	repeat_region	/rpt_family="Alu-Sp"	complement(30782..	.31066)					
	repeat_region	/rpt_family="LlMA10"	complement(31067..	.31170)					
	repeat_region	/rpt_family="Ll"	complement(31174..	.31463)					
	repeat_region	/rpt_family="Alu-Sz"	complement(31464..	.32119)					
	repeat_region	/rpt_family="Ll"	complement(32156..	.32238)					
	repeat_region	/rpt_family="LlMA2"	32253..	.32537					
	repeat_region	/rpt_family="Alu-Y"	complement(32539..	.33095)					
	repeat_region	/rpt_family="LlMA2"	33193..	.33469					
	repeat_region	/rpt_family="Alu-Jo"	complement(33485..	.33809)					
	repeat_region	/rpt_family="MER42C"	complement(33812..	.34098)					
	repeat_region	/rpt_family="Alu-Sg"							
	Query Match								
	Best Local Similarity								
	Matches 3062;								
	Conservative								
	Mismatches								
	Indels								
	Gaps								
QY	3066 GTGGATGAAATACATCTCAACTGTGAGGCACATTGCCATGAACCTTCCTTGTTCCCATCATCACT	3125							
Db	67898 GGTGGGATCGGCTCGACGTGCAGAGGGTTCGTGTCAGTAATTAGTAGGGATTCGCCACCTCTGTC	67839							
QY	3126 CGAAGACCCCCCATCAAGGGAGACCTCACGGGTGACACTTCGCCAGATGCCAGATGCCCTCAG	3185							
Db	67838 TCTCTCTCCCCTTCCAGGTGACACTCAGGTGACATTCGATGCCAGATGCCAGATGCCCTCAG	67779							
QY	3186 GCCCTGTCAATGATGACTCTCCACATTCACAGAGCGGACACAGAATGCATCAACTTCTTGCG	3245							
Db	67778 GCCCTGTCAATGATGACTCTCCACATTCACAGAGCGGACACAGAATGCATCAACTTCTTGCG	67719							
QY	3246 AAGGTATACGGCTACATGCCCCCTCTGTACACGACGATTCAGGGGTGANTCTGTGCTTTC	3305							
Db	67718 AAGGTATACGGGTACATGCCCCCTCTGTACACGAGTTTCAGGGGTGANTCTGTGCTTTC	67659							
QY	3306 AAGACACGGCTCGCCCATACACAAGAACCAAGTCTTCAAGTTATCTTAGGGGACAGACAG	3365							
Db	67658 AAGACACGGCTCGCCCATACACAAGAACCAAGTCTTCAAGTTATCTTAGGGGACAGACAG	67599							
QY	3366 GTCCTGGGAAGAAGATGAGCAGAGGAGAGAMAATGGCTCCCAAAGTTCTCTAGGACATTGCA	3425							
Db	67598 GTCCTGGGAAGAAGATGAGCAGAGGAGAGAMAATGGCTCCCAAAGTTCTCTAGGACATTGCA	67539							
QY	3426 GGACCTTGGGACATCTGCTGGGTGGGTGGCCAGAGCCTCTGCTGGAAGGGGACAGACAGA	3485							
Db	67538 GGACCTTGGGACATCTGCTGGGTGGGTGGCCAGAGCCTCTGCTGGAAGGGGACAGACAGA	67479							
QY	3486 GAAGTGAAGAAGAAACCCTGCTTATATCTTGAAGTCAACCAOACTGGCCCTGGAACCCCTG	3545							
Db	67478 GAAGTGAAGAAGAAACCCTGCTTATATCTTGAAGTCAACCAOACTGGCCCTGGAACCCCTG	67419							
QY	3546 GGCGGAGTCCCGGGGTTCCCGACAGCAGGGGACCTGATGATGATGATGATGATGATGATGATGAT	3605							

|||||
Db 67418 GGGGGAGTCCCCGGGGTCCCCACACAGGGCAGCTGATTTACCTTACACATGTGAGGACTGT 67359
OY 3606 GGGGACCTGACAGAGTACACACACGGTTGTCAGCCAGAGAGCGTGGTGGGTTT 3665
Db 67358 GGGGACCTGACAGAGTACACACGGTTGTCAGCCAGAGAGCGTGGTGGGTTT 67299
OY 3666 TACATTCATATACACATATATATATTTAAAAAGAAAGTTTCAGATTTCCATTCA 3725
Db 67298 TACATTCATATACACATATATATATTTAAAAAGAAAGTTTCAGATTTCCATTCA 67239
OY 3726 AGGCTATTTATATATATGTTGTTGATATATTAATACATGCACACACTTGCATACATATAT 3785
Db 67238 AGGCTATTTATATATATGTTGTTGATATATTAATACATGCACACACTTGCATACATATAT 67179
OY 3786 ATTTTGGGCTGGGGAGTGTGAGTTTGGCTTTCTAAGGAGGAGCCGACAGCTCCTT 3845
Db 67178 ATTTTGGGCTGGGGAGTGTGAGTTTGGCTTTCTAAGGAGGAGCCGACAGCTCCTT 67119
OY 3846 TGTTCGTATTTCTGGCGAGATGGGTCCCTGTGTGTCAGTGGCTTATCTTAAAGAT 3905
Db 67118 TGTTCGTATTTCTGGCGAGATGGGTCCCTGTGTGTCAGTGGCTTATCTTAAAGAT 67059
OY 3906 CATCTCCGATCTCCCGAGGCGCATCTGTGTGCAGCAACCAAGAAAGGATGAACTTGGCC 3965
Db 67058 CATCTCCGATCTCCCGAGGCGCATCTGTGTGCAGCAACCAAGAAAGGATGAACTTGGCC 66999
OY 3966 CTCTTGGGGCTGGAGCAAGGTCTCTTACCTTTCTGTTGCCAGTAGCAACCTGT 4025
Db 66998 CTCTTGGGGCTGGAGCAAGGTCTCTTACCTTTCTGTTGCCAGTAGCAACCTGT 66939
OY 4026 AACTCAGATTTCTCTTCCCACTGAATCCCTGGAGCGCTGACCTGGTGGGCTGTTCAGC 4085
Db 66938 AACTCAGATTTCTCTTCCCACTGAATCCCTGGAGCGCTGACCTGGTGGGCTGTTCAGC 66879
OY 4086 TTCTCTGCTGGGGCGGAGGATTTTGGAGATTTATCTTAAAGGCGCTGGCTCCGT 4145
Db 66878 TTCTCTGCTGGGGCGGAGGATTTTGGAGATTTATCTTAAAGGCGCTGGCTCCGT 66819
OY 4146 ACTTATCCCTGCTCTCCCATTTCTCTTGTGTTGAGAGAAATGAGAAAGCAAGATGTA 4205
Db 66818 ACTTATCCCTGCTCTCCCATTTCTCTTGTGTTGAGAGAAATGAGAAAGCAAGATGTA 66759
OY 4206 GAAAGAAATAGGGGCTGAAGACGCCACTCCAGATGAGCTCTTCTATCTGCTCTTGT 4265
Db 66758 GAAAGAAATAGGGGCTGAAGACGCCACTCCAGATGAGCTCTTCTATCTGCTCTTGT 66699
OY 4266 GAAAGCAACGTGCTGGGCTCAAGGCGTTTCTAAGTGCCTTCTTGTGATTTGGAACAG 4325
Db 66698 GAAAGCAACGTGCTGGGCTCAAGGCGTTTCTAAGTGCCTTCTTGTGATTTGGAACAG 66639
OY 4326 AGATCAGCAGGTCACATCTGCTGTGTCGAAGTGGTTTGCAGCTCCTCTCC 4385
Db 66638 AGATCAGCAGGTCACATCTGCTGTGTCGAAGTGGTTTGCAGCTCCTCTCTCC 66579
OY 4386 CTATGTAGAGCAAGCCAGTGTCTTGCAGAAACCAACCCGCGTGGCCGGGAAGTTTAC 4445
Db 66578 CTATGTAGAGCAAGCCAGTGTCTTGCAGAAACCAACCCGCGTGGCCGGGAAGTTTAC 66519
OY 4446 AGCAAGCGGCTGCTGGGTAATTCCTTGTGTAATTCACCTTCCCCCGGCTGTGC 4505
Db 66518 AGCAAGCGGCTGCTGGGTAATTCCTTGTGTAATTCACCTTCCCCCGGCTGTGC 66459
OY 4506 TGGAGCCCATCTCTGTGTTATCTGTGGTTTTTGGACCCCTTAATTCAGCTTGGCTGAGG 4565
Db 66458 TGGAGCCCATCTCTGTGTTATCTGTGGTTTTTGGACCCCTTAATTCAGCTTGGCTGAGG 66399
OY 4566 ACTCCCGAGGTTTGTATGTGCTAGAAACAATGGAGAGCTGTGATTTGCTGTGAAGCTC 4625
Db 66398 ACTCCCGAGGTTTGTATGTGCTAGAAACAATGGAGAGCTGTGATTTGCTGTGAAGCTC 66339
OY 4626 ACATCCAGCTTGGAAATCTAACGGGCAATTCACAACCCGAGTTACCACTTTCCACTCCCTG 4685
|||||

Db 66338 ACATCCAGCCTTGGAATCTAACGGGCAATTCACAACCCGAGTTTCCACTCCCTG 66279
OY 4686 CTTAGGATTTCTTCCCTGGGCTGAACACTGAATAAGCTAATTTTTTGGTCAGGTTGC 4745
Db 66278 CTTAGGATTTCTTCCCTGGGCTGAACACTGAATAAGCTAATTTTTTGGTCAGGTTGC 66219
OY 4746 AGTAGGGGAACCTTAGAGAGGTTGAGTGGCATTTGTCAAGGATTTAGCCATGACGTGT 4805
Db 66218 AGTAGGGGAACCTTAGAGAGGTTGAGTGGCATTTGTCAAGGATTTAGCCATGACGTGT 66159
OY 4806 TCTTGAACCTTACTTCTGTGAAGTGGAGTGTGACCTGGAAGTTTTCAGCACTGAACAA 4865
Db 66158 TCTTGAACCTTACTTCTGTGAAGTGGAGTGTGACCTGGAAGTTTTCAGCACTGAACAA 66099
OY 4866 AAGCTAGGTTTGTCTGTCATGTCACATGCTTAAAGCCAGTTCCGTTCCCTAGACCT 4925
Db 66098 AAGCTAGGTTTGTCTGTCATGTCACATGCTTAAAGCCAGTTCCGTTCCCTAGACCT 66039
OY 4926 TGGCATCTGTGCTTCTATTTTGGAAATACGTTTCTCTGTACACCTGTACACGTG 4985
Db 66038 TGGCATCTGTGCTTCTATTTTGGAAATACGTTTCTCTGTACACCTGTACACGTG 65979
OY 4986 GGTCTCTTCAAGTACTGTTTGAAGCTGGGCTCTTTGTGTAGCTCCACCCACTGTA 5045
Db 65978 GGTCTCTTCAAGTACTGTTTGAAGCTGGGCTCTTTGTGTAGCTCCACCCACTGTA 65919
OY 5046 GGGCTAGCTGGGCTTAAAGGGAACCTCCCATTTGGCAACCGGACCGGCGGCCAGG 5105
Db 65918 GGGCTAGCTGGGCTTAAAGGGAACCTCCCATTTGGCAACCGGACCGGCGGCCAGG 65859
OY 5106 ACTGTGTTCCAAAGTTTCCCGGCGCCCAACCCAGATAGGCTGATGCCCTGTG 5165
Db 65858 ACTGTGTTCCAAAGTTTCCCGGCGCCCAACCCAGATAGGCTGATGCCCTGTG 65799
OY 5166 AGGCAAGTGGTATGTTCACAGAGTGGGGTTCAGACGCCCTTCTCAGAACTTTTAG 5225
Db 65798 AGGCAAGTGGTATGTTCACAGAGTGGGGTTCAGACGCCCTTCTCAGAACTTTTAG 65739
OY 5226 TTGGCTCTACGTGACCTGACTTGTATTCCTTTTGAAGAGCTTCTTCCCTGGGG 5285
Db 65738 TTGGCTCTACGTGACCTGACTTGTATTCCTTTTGAAGAGCTTCTTCCCTGGGG 65679
OY 5286 AGCCAAAGATGTGTGTGTGGGCTATTTGTGGCTGCTATTTCAATCCTGTTCTTAA 5345
Db 65678 AGCCAAAGATGTGTGTGTGGGCTATTTGTGGCTGCTATTTCAATCCTGTTCTTAA 65619
OY 5346 ATGTAGGAACCTCAGATCTGACTTCACTGAGTGTGAGCTGCGTGAAGCGGCGTGTG 5405
Db 65618 ATGTAGGAACCTCAGATCTGACTTCACTGAGTGTGAGCTGCGTGAAGCGGCGTGTG 65559
OY 5406 TGGGACCCCTTTTAGCGGGAACGTAGTGGGCGGTCTGTGTGTGAGGCCAGGGCC 5465
Db 65558 TGGGACCCCTTTTAGCGGGAACGTAGTGGGCGGTCTGTGTGTGAGGCCAGGGCC 65499
OY 5466 TTCTCCCTTATGTGAGGACGAGTGTGCGGCGCCGAATGTCACTGTGTGATTAAGAAGG 5525
Db 65498 TTCTCCCTTATGTGAGGACGAGTGTGCGGCGCCGAATGTCACTGTGTGATTAAGAAGG 65439
OY 5526 CTGAGTGTGTGACACCAAAACATGCGCAGGAGGAGGCTGTGGTGGCTCTCAACA 5585
Db 65438 CTGAGTGTGTGACACCAAAACATGCGCAGGAGGAGGCTGTGGTGGCTCTCTCAACA 65379
OY 5586 AGGACACCCCTCTTGTGACCCCTGAAGGAACACTGGCTTGAAGACTGCAGACAGGCTGTG 5645
Db 65378 AGGACACCCCTCTTGTGACCCCTGAAGGAACACTGGCTTGAAGACTGCAGACAGGCTGTG 65319
OY 5646 AGGGGACGCGCTCTCTAGGAGAGGACGAAGGTGGCCACAGTGTCACTGTAGGTGC 5705
Db 65318 AGGGGACGCGCTCTCTAGGAGAGGACGAAGGTGGCCACAGTGTCACTGTAGGTGC 65259
OY 5706 TTCTCAGCAGGGGAACCGCGGCACTGTGACTGCTTGTGAGATGGGAAGGCGGCCACA 5765
Db 65258 TTCTCAGCAGGGGAACCGCGGCACTGTGACTGCTTGTGAGATGGGAAGGCGGCCACA 65199
|||||

Db 197884 GCTGTCACACCCGATCCGCTCAGCTGGCTCAGCTTCAGCTCTTTGTATCTGTCT 197943
QY 718 TCTTCCGCTCATCGCCACTATTACCTACACACTGTGATGAGGCTGATGAGGACGCA 777
Db 197944 TCTTCCGCTCATCGCCACTATTACCTACACACTGTGATGAGGCTGATGAGGACGCA 198003
QY 778 AGCGGATTTTGGTCCCCGGGTGGGGAACGAGCTGTGCGAGGTGAACGACGTCTGATC 837
Db 198004 AGCGGATTTTGGTCCCCGGGTGGGGAACGAGCTGTGCGAGGTGAACGACGTCTGATC 198063
QY 838 TGTGCGCATCCGGGAGTGGGTGATGAAGAGCTCCGACGTGGAGGCCAAGGCCAAG 897
Db 198064 TGTGCGCATCCGGGAGTGGGTGATGAAGAGCTCCGACGTGGAGGCCAAGGCCAAG 198123
QY 898 AGCTGAACAGCAGATCGCCAGCTGATCTGAAGATCGAAGCTGTAAAGAGCATTTG 957
Db 198124 AGCTGAACAGCAGATCGCCAGCTGATCTGAAGATCGAAGCTGTAAAGAGCATTTG 198183
QY 958 AGAAGCCAAAGCAGAGACCTGCTCAGCTCAAGAAATGTCATCAGCCAGACCGACATTTCT 1017
198184 AGAAGCCAAAGCAGAGACCTGCTCAGCTCAAGAAATGTCATCAGCCAGACCGACATTTCT 198243
QY 1018 ACAAGAGCTCATGGGCCAAGACAGGCCAAGCTGTCGCCATCCGACCTGCCAG 1077
Db 198244 ACAAGAGCTCATGGGCCAAGACAGGCCAAGCTGTCGCCATCCGACCTGCCAG 198303
QY 1078 AGAAGCAGATCGCGGCTCCTCCCGAAGGCCACTCGGGCTGCGGCTACACAACT 1137
Db 198304 AGAAGCAGATCGCGGCTCCTCCCGAAGGCCACTCGGGCTGCGGCTACACAACT 198363
QY 1138 GCTTTGATTTATTTCTCGTGGCTCTCAGCTCTGGCTTCCGGTCTACGTCTATGACAGTG 1197
Db 198364 GCTTTGATTTATTTCTCGTGGCTCTCAGCTCTGGCTTCCGGTCTACGTCTATGACAGTG 198423
QY 1198 ACCAGTTTCTTTGGAGCTACCTGGATCCCTTGGTCAAGCAGAGGCTTTCAAGGCGAG 1257
Db 198424 ACCAGTTTCTTTGGAGCTACCTGGATCCCTTGGTCAAGCAGAGGCTTTCAAGGCGAG 198483
QY 1258 CAGAGCTAACGTTTATGTTACAGAAATGCAAGACATCGCTCCTTTACGTATAGTAG 1317
Db 198484 CAGAGCTAACGTTTATGTTACAGAAATGCAAGACATCGCTCCTTTACGTATAGTAG 198543
QY 1318 TGGGAGAGATGCAAGAGCCGCTGCTGCGGCTGCTGAGCTGGAGAGCAAGTTGATT 1377
Db 198544 TGGGAGAGATGCAAGAGCCGCTGCTGCGGCTGCTGAGCTGGAGAGCAAGTTGATT 198603
QY 1378 CCGTGGCAGCTGGGAGCGGATGAGACACAACATGTCATCATCTGTCAAGTAACT 1437
198604 CCGTGGCAGCTGGGAGCGGATGAGACACAACATGTCATCATCTGTCAAGTAACT 198663
QY 1438 CAGATACACGAACCTTCTATAGCTACGTAGCTGCGCTGCTGATGGCCAGTCCA 1497
Db 198664 CAGATACACGAACCTTCTATAGCTACGTAGCTGCGCTGCTGATGGCCAGTCCA 198723
QY 1498 CATTACACACTGTCAGTACAGACCTGGCTTGAATTTGGTCTGATACACCGCTGTCCATG 1557
Db 198724 CATTACACACTGTCAGTACAGACCTGGCTTGAATTTGGTCTGATACACCGCTGTCCATG 198783
QY 1558 CCATGTCGAGCCCAATTCATGAAATCCACACAGTGGCGGTTGAACCGGAAATATC 1617
Db 198784 CCATGTCGAGCCCAATTCATGAAATCCACACAGTGGCGGTTGAACCGGAAATATC 198843
QY 1618 TCTTACACTTCCAGGGGAGAGATGAGTCTGTAGAGTCTAGCCCTTCCAGGAGCCGCT 1677
Db 198844 TCTTACACTTCCAGGGGAGAGATGAGTCTGTAGAGTCTAGCCCTTCCAGGAGCCGCT 198903
QY 1678 CATTGGAAGGAAATGAGGAGGCGCTCCCGGACTCGATGATGATCCGATATTGGCA 1737
Db 198904 CATTGGAAGGAAATGAGGAGGCGCTCCCGGACTCGATGATGATCCGATATTGGCA 198963
QY 1738 CCCTGAAGGCGGTGACAGAGCAGCAGCTGATAGGTCTGTGGAATTAACCTGCAAAA 1797
Db 198964 CCCTGAAGGCGGTGACAGAGCAGCAGCTGATAGGTCTGTGGAATTAACCTGCAAAA 199023

QY 1798 ACCAGCCAAACCAGACCTGCGACTGAGTGGGCACTGTGTGAGAGCGGAGAGCCGCT 1857
Db 199024 ACCAGCCAAACCAGACCTGCGACTGAGTGGGCACTGTGTGAGAGCGGAGAGCCGCT 199083
QY 1858 TGAATTTGCTGAAGCTCTCCACTTGGCCCTCATTTATTTACCCCCGGGAGACCTCGCTTGG 1917
Db 199084 TGAATTTGCTGAAGCTCTCCACTTGGCCCTCATTTATTTACCCCCGGGAGACCTCGCTTGG 199143
QY 1918 TTAATTCCTGTGGGTGTGCAACAGGCTCTTGAACCCCGGAAAGTCGGTGCCTCCGG 1977
Db 199144 TTAATTCCTGTGGGTGTGCAACAGGCTCTTGAACCCCGGAAAGTCGGTGCCTCCGG 199203
QY 1978 TGGTGTGGGGGAGCAGGTCCAGCTTCCCTACAGAGACATGCTGAGTGGAAAGAGGCGG 2037
Db 199204 TGGTGTGGGGGAGCAGGTCCAGCTTCCCTACAGAGACATGCTGAGTGGAAAGAGGCGG 199263
QY 2038 CCTGTGTGTCCAAAGCCTCGTGTACCGAGGTTATTTCTGCTCAGAGCCTCTCGG 2097
Db 199264 CCTGTGTGTCCAAAGCCTCGTGTACCGAGGTTATTTCTGCTCAGAGCCTCTCGG 199323
QY 2098 ATAGTACCTCTGCGCTATGAGGCGGCAAGGCGCTTCTGTGGAGACTTACTTCTCA 2157
Db 199324 ATAGTACCTCTGCGCTATGAGGCGGCAAGGCGCTTCTGTGGAGACTTACTTCTCA 199383
QY 2158 CTGCTGACAGTATTTTAAATACCGTCTGAGCTATGATTAAGAGTCCGATCCAGTCCAG 2217
Db 199384 CTGCTGACAGTATTTTAAATACCGTCTGAGCTATGATTAAGAGTCCGATCCAGTCCAG 199443
QY 2218 CGGCTCCCATCCGGGAAGAGCGGCGAGCTGAGATCCCCACCGTTACAGCAGGCGCTG 2277
Db 199444 CGGCTCCCATCCGGGAAGAGCGGCGAGCTGAGATCCCCACCGTTACAGCAGGCGCTG 199503
QY 2278 GAACTGACCCCAATGAGCTGACACAGGAGACCTGGACCTGGAGGCGGAGTGGAGACGAGC 2337
Db 199504 GAACTGACCCCAATGAGCTGACACAGGAGACCTGGACCTGGAGGCGGAGTGGAGACGAGC 199563
QY 2338 CGCCCTACGCTTACCCAGATACCTCCGCAATTTCACTGTGACTGTACGTATTTAAC 2397
Db 199564 CGCCCTACGCTTACCCAGATACCTCCGCAATTTCACTGTGACTGTACGTATTTAAC 199623
QY 2398 GCAGCTGGAAGTGTGCTCCAGAGGCTTTTCCATTTTCCCCACACTCCCTTTGACCCG 2457
Db 199624 GCAGCTGGAAGTGTGCTCCAGAGGCTTTTCCATTTTCCCCACACTCCCTTTGACCCG 199683
QY 2458 TGTGCTCCAGAGGCAAAATTTTGGGCTCAGGAGTGGCTTTCGCGCTATTTGGTGTG 2517
Db 199684 TGTGCTCCAGAGGCAAAATTTTGGGCTCAGGAGTGGCTTTCGCGCTATTTGGTGTG 199743
QY 2518 GAGCTGGGGTCTGCGAAGAAATTTTCAGGACGCTTGGAGGCAATGTTCCCGAGAGC 2577
Db 199744 GAGCTGGGGTCTGCGAAGAAATTTTCAGGACGCTTGGAGGCAATGTTCCCGAGAGC 199803
QY 2578 AGTTACGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2637
Db 199804 AGTTACGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 199863
QY 2638 GGCTGATGAGCTTCCCTTACCTGGAACAGTCTGTGTGTGATGATGATGATGATGATGATGATGAT 2686
Db 199864 GGCTGATGAGCTTCCCTTACCTGGAACAGTCTGTGTGTGATGATGATGATGATGATGATGATGAT 199912

RESULT 12
AF083550 3479 bp mRNA linear ROD 02-JUL-1999
LOCUS AF083550 Mus musculus exostotin EX111 mRNA, complete cds.
DEFINITION AF083550
ACCESSION AF083550
VERSION AF083550.1 GI:5326768
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1720 ATGACCGGATATTGGCCACCTGAAAGCGGTGCAGAGACGACGATGATGATCTCTGG 1779
1581 AGCATTCGATATTCGCCACCTTAAAGGCTGTACAGACAGACGATGATGATGATGCTGG 1640
1780 TGGAAATTCACCTGCAAAAACAGCCCAAAACCCAGCTGCGGACGAGAGGCGGACATGCTG 1839
1641 TAGAAATTCACCTGCAAAAACAGCCCAAAACCCAGCTGCGGACGAGAGGCGGACATGCTG 1700
1840 GAGAGCGGAGGAGACCGCTTGGAAATTCGTGAAGCTCTCCACCTTCCGCTCATCTATCC 1899
1701 GGGAGCGGAGGAGACCGCTTGGAAATTCGTGAAGCTCTCCACCTTCCGCTCATCTATCC 1760
1900 CCGGAGACCGCTTGGCTTGTATTTCTCTGCGGTGTCACACAGCGCTCTTGAAGCGCTGG 1959
1761 CCGGAGACCGCTTGGCTTGTATTTCTCTGCGGTGTCACACAGCGCTCTTGAAGCGCTGG 1820
1960 AAGTCGGGTGCGTCCGGTGTGCTGGGGGAGAGGTCGACCTTCCGATACGAGGACATTC 2019
1821 AGGTGGGGGCGCTGCGGTGTGCTGGGGGAGAGGTCGACCTTCCGATACGAGGACATTC 1880
2020 TGCAGTGGAAAGAGGCGGCGCTGGTGTGTCGCAAGCGCTCTGTATACGAGGTTCAATTC 2079
1881 TGCAGTGGAAAGAGGCGGCGCTGGTGTGTCGCAAGCGCTCTGTATACGAGGTTCAATTC 1940
2080 TGCACAGAGCGCTCTCGGATATGATACCTCTGCTATGAGCGGCAAGCGCTTCTCT 2139
1941 TGTACAGAGCGCTCTCGGATATGATACCTCTGCTATGAGCGGCAAGCGCTTCTCT 2000
2140 GGGAGACTTACTTCCACCTCTGACAGTATTTTAAATACGCTGCTGCTGATGATGGA 2139
2001 GGGAGACTTACTTCTCCACCGGACAGAGATTTTAAATACGCTGCTGCTGATGATGGA 2060
2200 CTGCGATTCAGATCCGACCGCTCCATCCGGAAGAGGCGGACGATGATGATGATGGA 2259
2061 CTGCGATTCAGATCCGACCGCTCCATCCGGAAGAGGATGATGATGATGATGATGGA 2120
2260 GTTACAGGAGGCGGCTGGAGATGATGATGATGATGATGATGATGATGATGATGATGGA 2319
2121 GTTACAGGAGGCGGCTGGAGATGATGATGATGATGATGATGATGATGATGATGATGGA 2180
2320 GGGCAGTGGAGAGGAGGCGGCTGACGCTACGCTACGCTACGCTACGCTACGCTACGCTGA 2379
2181 GGGCAGTGGAGAGGAGGCGGCTGACGCTACGCTACGCTACGCTACGCTACGCTACGCTGA 2240
2380 CTGTCACGATCTTTTACCGGACGCTGGAATGCTCCAGGCGCTTTCATCTTTTCCCC 2439
2241 CTGTCACGATCTTTTACCGGACGCTGGAATGCTCCAGGCGCTTTCATCTTTTCCCC 2300
2440 ACATCTCCCTTGGACCTGTGTGCTGCTGACAGGCGCAATTTTGGGCTCAGGAGCTGCT 2499
2301 ACATCTCCCTTGGACCTGTGTGCTGCTGACAGGCGCAATTTTGGGCTCAGGAGCTGCT 2360
2500 TTTGGGCTTGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAG 2559
2361 TTTGGGCTTGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAG 2420
2560 GCAATGTTTCCCGAGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAG 2619
2421 GCAATGTTTCCCGAGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCTGAG 2480
2620 TTTAGAACTCTTTAGAGAGGCTGAGGAGGCTTCTTACCTTAACAAGGCTGCTGCTGCT 2679
2481 TCAATGAACTCTTTAGAGAGGCTGAGGAGGCTTCTTACCTTAACAAGGCTGCTGCTGCT 2540
2680 GGAATTTTCCCGAGAGGCTGAGGAGGCTTCTTACCTTAACAAGGCTGCTGCTGCTGCT 2739
2541 GGAATTTTCCCGAGAGGCTGAGGAGGCTTCTTACCTTAACAAGGCTGCTGCTGCTGCT 2600
2740 TGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2739
2601 TGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2660

2800 AGACAGAGGCGCATCTCTCCATTTGATGACATGCTCACTCCGCGCATGACGAATTCATGT 2859
2661 AGACAGAGGCGCATCTCTCCATTTGATGACATGCTCACTCCGCGCATGACGAATTCATGT 2720
2860 TTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2919
2721 TTTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2780
2920 AGCAGTGGAGCATCTCCCATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2979
2781 ATGCTGGAGCATCTCCGACACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2840
2980 CCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3039
2841 CCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2900
3040 TGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3099
2901 TGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2960
3100 TGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3159
2961 TGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3020
3160 CATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3219
3021 CATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3080
3220 GGCACAGTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3279
3081 GGCACAGTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3140
3280 AGTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3339
3141 AGTTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3200
3340 TCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3399
3201 TCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3259
3400 GGCCTCCCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3459
3260 ACCCTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3318
3460 AGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3518
3319 ACCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3378
3519 GTCAGCCACACTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3578
3379 GTCAGCCACACTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3436
3579 TGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3605
3437 TGAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3463

RESULT 13
AB033367 3198 bp mRNA linear ROD 14-Apr-2000
LOCUS AB033367
DEFINITION Rattus norvegicus Regr mRNA for Reg receptor, complete cds.
ACCESSION AB033367
VERSION AB033367.1 GI:7339683
KEYWORDS Reg receptor.
SOURCE Rattus norvegicus (strain:Wistar) adult male Islets of Langerhans
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (sites)
Kobayashi, S., Akiyama, T., Nata, K., Abe, M., Teijima, M.,
Shervani, N.J., Umno, M., Matsuno, S., Sasaki, H., Takasawa, S. and

TITLE Okamoto, H.
Identification of a receptor for reg (regenerating gene) protein, a
pancreatic beta-cell regeneration factor
JOURNAL J. Biol. Chem. 275 (15), 10723-10726 (2000)
MEDLINE 20219101
REFERENCE 2 (bases 1 to 3198)
AUTHORS Okamoto, H. and Takasawa, S.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1999) Shln Takasawa, Tohoku University Graduate
School of Medicine, Department of Biochemistry, 2-1 Seiry-machi,
Aoba-ku, Sendai, Miyagi 980-8575, Japan
(E-mail:shln@swemail.cc.tohoku.ac.jp, Tel:+81-22-717-8080,
Fax:+81-22-717-8083)

FEATURES
Source Location/Qualifiers
1..3198
/organism="Rattus norvegicus"
/strain="Wistar"
/db_xref="taxon:10116"
/sex="male"
/tissue_type="Islets of Langerhans"
/dev_stage="adult"
5..2764
/gene="Reg"
5..2764
/gene="Reg"
/codon_start=1
/product="Reg receptor"
/protein_id="BA092895.1"
/db_xref="GI:7339684"
/translation="MTCGTLRANGVNGGOTCMRNSRIRLTWLSFTLIIIVFP
LIAHYLTLDADBAKRTIRPRAGNLCVKKVLDICRIRESVSELDLEAKROE
LNSIAKLNLKTECKKSLENKODLLOLKNVISOEHSYELAAQNPRLSLRL
PERKDAGLPRPKVGRGRLHNKCPDYSRCPDSEVYDSDOFAGYIDPLKQAF
QATVRAVYVYENAIACLYVYVGEIOEPVLDPADLEKOLSPHMRDGNHVI
NLSKSDTONLNVSTGRAMVAOSTFYAAOYRAGPOLYSPYHMASSEPMETIPQ
VPRKYLTFEGEKISLSIOEARSPEEMGDDPADDDNIATLVAQDSKID
OIVLEPTCKNPKRSLPTENALCGERDRLELTLSTFALLITGDESLIISGCAR
LFEALVGAIPVYVGEVQVLPYHDMLOMNEALVPRRYTEVHFLRLSDSLAM
RROGRFLMETEYFSTADSIIFNTVLAIMRTIOIPAPVREYAAAFPHRSGAAGTDN
MADNGDLDPVETEPYASPKYLRFNTLTVDYRSMNASPGFHLPPRPDPVLP
SEAKFLSGTGFPIGGAGSGKEFOALAGVORROFPYVMTTIRREYLNKSLER
LNGLPYLNKVVVWNSPKLPSEDLMPDIVPIVYVTEKNSLNNRFLPNNELTEAI
LSDDDAHRLHDELMGFRVYRERDRIVGPGRYHAMDI PHOSWLNYSVSELSAV
LTSADPHKRYATLYSTVMPQALRDVDEYINCEDIAMNPLVSHITKPKIKYTSMT
FRCPGCPQALSHDSHHERHKCINFEKVYGYMPLLYTOFRVDSVLFKRLHDKRX
CFKFI"

BASE COUNT 723 a 882 c 850 g 743 t
ORIGIN

Query Match 38.0% Score 2347.4; DB 10; Length 3198;
Best Local Similarity 85.3%; Pred. No. 0;
Matches 2721; Conservative 0; Mismatches 416; Indels 54; Gaps 7;

QY 590 ACATCATGAGGCGTATACATGCTGGCGGAGTGGGGGCGGGGAACGAGAGTCAGACCTG 649
Db 1 ACTCATGACAGGCGTATACATGCTGGCGGAGTGGGGGAGTGGGAACGCTGACAGACCTG 60
QY 650 CATGCTGCGCTGTGCAACCGCATCGCTCAGCTGAGCTAGGCTTACAGCTCTTTGTCAT 709
Db 61 TATGCTGCGCTGTGCAACCGCATCGCTGAGCTAGGCTTACAGCTCTTTGTCATCAT 120
QY 710 CCGGCTCTTCTCCGCTCATGCGCCACTATTAATCTACACACTCTGGATGAGGCAATGA 769
Db 121 CCGGCTCTTCTCCGCTCATGCGCCACTATTAATCTACACACTCTGGATGAGGCAATGA 180
QY 770 GCGAGCAAGCGGATTTTGGTCCCGGGTGGGAACGAGCTGTGCGAGGTGAAGACAGT 829
Db 181 GCGCGCAAGCGGATTTTGGTCCCGGGTGGGAACGAGCTGTGCGAGGTGAAGACAGT 240
QY 830 GCTGATCTGTGCGGATCGGGAGTGTGAGTGAAGAGCTCTGAGGCTGAGGACCA 889
Db 241 CCTAGATCTTGTGCGGATCGGGAGTGTGAGTGAAGAGCTTCTTACAGCTGAGGACCA 300

QY 890 GCGCAAGAGCTGATACAGCGAGATGCCAAGCTGAATCTGAAGTCAAGCTGTAGAA 949
Db 301 GCGCAGAGAGCTGAAACAGCAGATTTGCCAAGCTAAACCTGAAGTTAACCTGTAGAA 360
QY 950 GAGATGTAGAGAGCGCAAGAGAGAGCTGCTCCAGCTCAAGATGTCAACGCCAGACCA 1009
Db 361 GAGTATAGAGAGCGCAAGAGAGAGCTGCTCCAGCTCAAGATGTCAAGATGTCAAGCA 420
QY 1010 GCATTCCTCAAGAGAGCTGATGCGCCAGAACCAAGCCCAAGCTGCTGCTGCTGCT 1069
Db 421 GCATTCCTCAAGAGAGCTGATGCGCCAGAACCAAGCCCAAGCTGCTGCTGCTGCTGCT 480
QY 1070 GCTCCAGAGAGAGAGAGATGCGCGCTCCCTCCCGAAGCCCAAGCGGAGCTGCTGCT 1129
Db 481 GCTCCAGAGAGAGAGAGATGCGCGCTCCCTCCCGAAGCCCAAGCGGAGCTGCTGCTGCT 540
QY 1130 ACACAGCTGCTTGTATATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1189
Db 541 ACACAGCTGCTTGTATATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
QY 1190 TGACAGTGAACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1249
Db 601 TGACAGTGAACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
QY 1250 GCGCAGAGAGAGAGCTTACGTTATGTACAGAAATGACAGACATCGCTGCTTACGT 1309
Db 661 GCGTACAGAGAGAGAGCTTACGTTATGTACAGAAATGACAGACATCGCTGCTTACGT 720
QY 1310 GATCTAGTGGAGAGAGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1369
Db 721 GGTGTATGAGAGAGAGATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 1370 GTTGTATGAG 1429
Db 781 GGTGTATGAG 840
QY 1430 AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1489
Db 841 CCGGAG 900
QY 1490 CCGAG 1549
Db 901 CCGAG 960
QY 1550 GGTCCATGAG 1609
Db 961 GGTCCATGAG 1020
QY 1610 GAATATCTGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1669
Db 1021 GAATATCTGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1080
QY 1670 GGGCGGCTCTTCAAG 1729
Db 1081 GGGCGGCTCTTCAAG 1140
QY 1730 CATGCGACCGCTGAAGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1789
Db 1141 CATGCGACCGCTGAAGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1790 CTGCAAAAACAGCCCAAAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1849
Db 1201 CTGCAAAAACAGCCCAAAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
QY 1850 GGAACGCTTGAATTTGCTGAGAGCTGCAACCTGCGCTCATCATTAACCCCGGGAGCC 1909
Db 1261 GGAACGCTTGAATTTGCTGAGAGCTGCAACCTGCGCTCATCATTAACCCCGGGAGCC 1320
QY 1910 TCGCTTGTATATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1969
Db 1321 GAGCGTGTATATCTATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380

QY 1970 CGTCCGCTGCTGCTGGGAGACAGTCCAGCTTCCCTACCAGACATGCTGCAGTGGAA 2029
 Db 1381 TGTCTCTGCTCTCTCTGGGAGAGAGTGCAGTCCCTGACACAGTGCATACATGAA 1440
 QY 2030 CGAGGCGGCGCTGCTGGTGGCCAAAGCCTGGTACAGAGTTCATTTCCGACAGAG 2089
 Db 1441 TGAAGCGCGCTGCTGGTGGCCAAAGCCTGGTACAGAGTTCATTTCCGACAGAG 1500
 QY 2090 CTTCTCCGATAGTACCTCTCTGCTATAGAGCGGCAAGCGCGCTTCTCTGGAGACTTA 2149
 Db 1501 TCTGTACAGAGATGATGTTGGCTATAGAGCGGCAAGCGCGCTTCTCTGGAGACTTA 1560
 QY 2150 CTTTCCAGTCTGACAGATATTTTAAATACCTGCTGGCTATATAGAGTCCGATCCA 2209
 Db 1561 CTTTCCAGTCTGACAGATATTTTAAATACCTGCTGGCTATATAGAGTCCGATCCA 1620
 QY 2210 GATCCAGCGCTGCTCCATTCGGGAGAGGCGGAGCTGAGATCCCGACGCTTGGAGCAA 2269
 Db 1621 GATCCAGCTGCTCTCCATTCGGGAGAGGCTGAGATCCCGACGCTTGGAGCAA 1680
 QY 2270 GCGGCTGGAACCTGACCCCAACATGCTGACACAGGAGGAGCTGAGGAGGCGGAGTGA 2329
 Db 1681 GCGAGCTGCTGCTGACCCCAACATGCTGACACATGAGGAGGAGCTGAGGAGGCGGAGTGA 1740
 QY 2330 GACGAGAGCGGCTTACGCTTACCCAGATACCTCCGCAATTTCTGACTGTACTGA 2389
 Db 1741 GACGAGAGCGGCTTACGCTTACCCAGATACCTCCGCAATTTCTGACTGTACTGA 1800
 QY 2390 CTTTACCGGAGTGAACGCTGCTGCTGAGGAGGCTTTCATCTTTCCCGGACATCCCTT 2449
 Db 1801 CTTGTTACCGGAGTGAACGCTGCTGCTGAGGAGGCTTTCATCTTTTCCACACACACCTT 1860
 QY 2450 TGACCTGTGTGCTTCCCTGAGAGGCAATTTCTGGGCTCAGGAGCTGGCTTCCGCTAT 2509
 Db 1861 TGACCTGTGTGCTTCCCTGAGAGGCAATTTCTGGGCTCAGGAGCTGGATTTCCGCTAT 1920
 QY 2510 TGTGTGTGAGTCTGGGCTTCTGGCAAGAAATTTTCAGGAGGCTTGGAGGCAATTTCC 2569
 Db 1921 CCGTGTGTGAGTCTGGGCTTCTGGCAAGAAATTTTCAGGAGGCTTGGAGGCAATTTCC 1980
 QY 2570 CCGAGAGAGTTCAGGAGGCTGATGTGATGATGATGATGATGATGATGATGATGATG 2629
 Db 1981 GCGGAGAGAGTTCAGGAGGCTGATGTGATGATGATGATGATGATGATGATGATGATG 2040
 QY 2630 TTTAGAGAGGCTGAAATGGCTTCCCTACCTGACACAGGCTGCTGCTGCTGCTGCTG 2689
 Db 2041 CCTGAGAGGCTGAAATGGCTTCCCTACCTGACACAGGCTGCTGCTGCTGCTGCTGCTG 2100
 QY 2690 CAAGCTGCTCAGAGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2749
 Db 2101 CAAGCTGCTCAGAGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
 QY 2750 TACTGAGAGAGAGTTCGAACAGGATTTCTTACCCGAGAAATGGAATGAGAGAGG 2809
 Db 2161 TACTGAGAGAGAGTTCGAACAGGATTTCTTACCCGAGAAATGGAATGAGAGAGG 2220
 QY 2810 CATCTCTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2869
 Db 2221 AATATTTGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2280
 QY 2870 GGTGTGAGAGAGGCTGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2929
 Db 2281 GGTGTGAGAGAGGCTGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2340
 QY 2930 CATCCCCCATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2989
 Db 2341 CATCCCCCATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2400
 QY 2990 GACAGTGTGCTCTTCTTCAAGATATATATATATATATATATATATATATATATATAT 3049
 Db 2401 GACAGTGTGCTCTTCTTCAAGATATATATATATATATATATATATATATATATATAT 2460
 QY 3050 GCGCATCGGAGACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3109

Db 2461 GCGCATCGGAGACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
 QY 3110 TGTCTCCATCATCATCTGGAAGCGGCGGCTTACAGTCTGCTGCTGCTGCTGCTGCTGCT 3169
 Db 2521 TGTCTCCATCATCATCTGGAAGCGGCGGCTTACAGTCTGCTGCTGCTGCTGCTGCTG 2580
 QY 3170 CCGAGATGCGCTTCAAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3229
 Db 2581 CCGAGATGCGGCTTCAAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2640
 QY 3230 CATCAATCTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 3289
 Db 2641 TATCAATCTTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2700
 QY 3290 GATTTGTGCTCTTCAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3349
 Db 2701 GATTTGTGCTCTTCAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2760
 QY 3350 GATTTGTGCTCTTCAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3408
 Db 2761 GATTTGTGCTCTTCAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2820
 QY 3409 GATTTGTGCTCTTCAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3467
 Db 2821 GATTTGTGCTCTTCAAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2877
 QY 3468 CTTGGAAGGAG 3527
 Db 2878 CTTGGAAG 2936
 QY 3528 ACTGAGGCTGAGAGGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3587
 Db 2937 ACTGAGGCTGAGAGGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2982
 QY 3588 GCTTACAGTCTGAGAGGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3647
 Db 2983 GCTTACAGTCTGAGAGGCTGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3014
 QY 3648 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3707
 Db 3015 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3068
 QY 3708 TTTGAGATTTGCTTCAAGGCTTATTTATATATATATATATATATATATATATATAT 3767
 Db 3069 TTTGAGATTTGCTTCAAGGCTTATTTATATATATATATATATATATATATATATAT 3128
 QY 3768 ACACTTGCTAT 3778
 Db 3129 ATACATGTATA 3139
 RESULT 14
 AC087099/c 236447 bp DNA linear HTG 24-FEB-2001
 LOCUS Mus musculus clone RP23-401A5, *** SEQUENCING IN PROGRESS *** 44
 DEFINITION Mus musculus clone RP23-401A5, *** SEQUENCING IN PROGRESS *** 44
 ACCESSION AC087099
 VERSION AC087099.3 GI:13122753
 KEYWORDS HTG, HTGS, PHASE1.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 236447)
 AUTHORS McCombie, W.R., Baker, J.P., Bahret, A., Bal, H., Ballia, V.,
 Bedhia, N.N., de la Bastide, M., Huang, E.N., King, L., Kirchoff, K.A.,
 Miller, B., Nascimento, L.O., O'Shaughnessy, A.L., Preston, R.R.,
 Rodriguez, S., Santos, L., Shah, R.S., Splegl, L.A., Toth, K., Vail, M.D.
 and Zutter, T.
 TITLE Mouse Genomic Sequence
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 236447)

AUTHORS
TITLE
JOURNAL

COMMENT

McCombie, W.R.
Direct Submission
Submitted (07-Dec-2000) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
On Feb 24, 2001 this sequence version replaced gi:12658024.
----- Genome Center

Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory

Center code: CSHL

Web site: <http://www.cshl.org/genseq>

Contact: mccombie@cshl.org

----- Project Information

Center project name: RP23-401A5

Center clone name: RP23-401A5

* NOTE: This is a 'working draft' sequence. It currently
* consists of 44 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 14598: contig of 14598 bp in length
* 14599 14681: gap of unknown length
* 14682 28334: contig of 13653 bp in length
* 28335 28417: gap of unknown length
* 28418 38223: contig of 10806 bp in length
* 38224 39307: gap of unknown length
* 39307 49565: contig of 10259 bp in length
* 49565 49649: gap of unknown length
* 49649 59905: contig of 10257 bp in length
* 59905 59989: gap of unknown length
* 59989 69333: contig of 9345 bp in length
* 69334 69416: gap of unknown length
* 69417 78200: contig of 8784 bp in length
* 78201 78283: gap of unknown length
* 78284 86729: contig of 8445 bp in length
* 86729 86812: gap of unknown length
* 86812 94700: contig of 7889 bp in length
* 94701 94783: gap of unknown length
* 94784 102340: contig of 7557 bp in length
* 102341 102423: gap of unknown length
* 102424 109259: contig of 6836 bp in length
* 109260 109342: gap of unknown length
* 109343 116029: contig of 6687 bp in length
* 116030 116112: gap of unknown length
* 116113 122726: contig of 6614 bp in length
* 122727 122809: gap of unknown length
* 122810 129419: contig of 6610 bp in length
* 129420 129502: gap of unknown length
* 129503 135897: contig of 6395 bp in length
* 135898 135980: gap of unknown length
* 135981 142114: contig of 6134 bp in length
* 142115 142197: gap of unknown length
* 142198 148219: contig of 6022 bp in length
* 148220 148302: gap of unknown length
* 148303 154085: contig of 5783 bp in length
* 154086 154168: gap of unknown length
* 154169 158668: contig of 4500 bp in length
* 158669 158751: gap of unknown length
* 158752 163218: contig of 4467 bp in length
* 163219 163300: gap of unknown length
* 163301 167543: contig of 4243 bp in length
* 167544 167625: gap of unknown length
* 167626 171801: contig of 4176 bp in length
* 171802 171883: gap of unknown length
* 171884 175948: contig of 4065 bp in length
* 175949 176030: gap of unknown length
* 176031 179961: contig of 3931 bp in length
* 179962 180043: gap of unknown length
* 180044 183950: contig of 3907 bp in length

```

```

* 183951 184032: gap of unknown length
* 184033 187884: contig of 3852 bp in length
* 187885 187966: gap of unknown length
* 187967 191816: contig of 3850 bp in length
* 191817 191898: gap of unknown length
* 191899 195331: contig of 3433 bp in length
* 195332 195413: gap of unknown length
* 195414 198711: contig of 3288 bp in length
* 198712 198793: gap of unknown length
* 198794 201984: contig of 3191 bp in length
* 201985 202066: gap of unknown length
* 202067 205007: contig of 2941 bp in length
* 205008 205089: gap of unknown length
* 205090 207988: contig of 2899 bp in length
* 207989 208070: gap of unknown length
* 208071 210947: contig of 2877 bp in length
* 210948 211029: gap of unknown length
* 211030 21859: contig of 2830 bp in length
* 21860 21941: gap of unknown length
* 21942 216606: contig of 2665 bp in length
* 216607 216688: gap of unknown length
* 216689 219134: contig of 2446 bp in length
* 219135 219216: gap of unknown length
* 219217 221605: contig of 2389 bp in length
* 221606 221687: gap of unknown length
* 221688 223857: contig of 2170 bp in length
* 223858 223939: gap of unknown length
* 223940 226000: contig of 2061 bp in length
* 226001 226082: gap of unknown length
* 226083 228132: contig of 2050 bp in length
* 228133 228214: gap of unknown length
* 228215 230236: contig of 2022 bp in length
* 230237 230318: gap of unknown length
* 230319 232337: contig of 2018 bp in length
* 232338 232418: gap of unknown length
* 232419 234397: contig of 1979 bp in length
* 234398 234479: gap of unknown length
* 234480 236365: contig of 1886 bp in length
* 236366 236447: gap of unknown length.

```

```

FEATURES
    source
        1..236447
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="RP23-401A5"

```

```

BASE COUNT      64055 a 51700 c 50924 g 65787 t 3981 others
ORIGIN

```

```

Query Match      17.58; Score 1077.8; DB 2; Length 236447;
Best Local Similarity 82.08; Pred. No. 4.1e-244;
Matches 1314; Conservative 0; Mismatches 257; Indels 32; Gaps 5;

```

```

QY 194 TAGACACTATGCAATATTGTTGGTCACGAAACCAAGAAAGAGCTATGGCATTTGAA 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 236057 TCGAGAGTATGCAATATGGCTGCCGACAAAGCAAGGACAAAGCTATAGCCGTTAAA 235998
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 254 AAGTCTGTGTGATTCAGAGGTGTTTTTCTGGGTTTCATCATCAGGTACCTCTCCCT 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235997 ATGCTCCCTCGATGCAGAGGCTCTTCCCTGGCTTTAGCACCAATGATCTCTCTTT 235938
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 314 TCATTCAGCAAAAGATGTCACCTTTATGCTTTGATTAAGATTAAAGACATGTTCTTT 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235937 TCATTCATCAGCAAAAGATGTTGTTACTTCTTACTATGATGATG----- 235894
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 374 GGTCAACAGCCAGAACTTAATCTGCTGGAATAGGGTCAAGACCATTTACGTCGACG 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235895 -----ACAGCTGTACTTCAATTCAGATTTGCTGACGACGATGAACCTGATTCGCG---- 235843
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 434 TGAGAAAATGAATGTTCTATTTATTTGGTGCCTGTCTGGGAGACACACTAACTCTTC 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 235844 TGAGGAAGATGAATTTCCATTTTACTTGTGCTGCTTGTCAGGAGACACACTGATCTTTC 235785
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 494 TGGAAAAGCTGTGATGAGAAACAGAGATGTTTGTGGAATAGCAACCATGTTATGCGCA 553

```


D	235784	CAGAAACTGTGTGTGGAAA -AGAGTTTGGCGTTTGTGCAGACAGACATCATGTTATGGCGA	235726
OY	554	GTGACCCGACGTGATCTG --- GGGGGCAGGCTGCAGAGACTCATGACAGGCTATACAT	610
D	235725	GGGATCCCACTGTATCAGAGTGGGCAAGAGCACACGGCAACTATGCACAGGCTATACAT	235666
OY	611	GCTGCGAATATGGGGGCGGGGGAACGGAGGACAGACACTTGCACTGCGGGTGGTCAACCG	670
D	235665	GTTGGGAAATGGGGGAGTGGGGGAACGGTGGTCAAGACTGTATGTCTGGCTGTGTCAATGC	235606
OY	671	CATCCGCTCAGCTGGGCTCAGCTTTCACGCTTCTTGTATCTTCCTGGTCTTCTCCGCTCAT	730
D	235605	CATCCGGCTGTGACATGGGCTGAGATTTCAACGCTTGTATCATATCTCTGCTTCTTCCCTCAT	235466
OY	721	GGCCCACTAATTAACGTCAACCACTGTGATGAGGGGATGAGGACAGGCAAGCGGATTTTGTG	790
D	235545	TGCTTACTATTAACCTCACCACTCTGGACGAGGACGAGACAGAGGCTGGCAAGCGCATTTCCG	235486
OY	791	TCCCGGGGTGGGGAACGAGCTGTGCGAGGTGAAACAGCTGTGGATGTGTCCGATCCG	850
D	235485	CCCTCGGGCTGGAGTGAGCTCTGTGAGGTAAACACAGCTCTGATCTGTGTGGATTCG	235426
OY	851	GGAGTTCGTTGAGTGAAGAGCTTCCTGCAAGCTGTGAGGCGCCACAGAGCTGAACAGCA	910
D	235425	TGAGTCTGTGAGCAGAGAGCTTCTACAGCTGCAAGCCAGACCGGAGAGGCTGAACAGCGA	235366
OY	911	GATCCCAAGCTGAATCGAAGATCGAAGGCTGTATACAGAGCATTTAGAAACCCAAAGCA	970
D	235365	CATTGCCAAGCTGAAACCTCAAGATTTGAAGCTGTATACAGAGCATTTAGAAATGCCAAGCA	235306
OY	971	GGACCTGCTCCAGCTCAAGAAATGTATCAGCGACAGCCAGACATTTCTTCAAGAGCTCAT	1030
D	235305	GGACCTGCTGAGCTCAAGAAATGTATTAAGCCAGACAGACACTCTTCAAGAGGCTGAT	235246
OY	1031	GGCCCCAAGAACGAGCCAAAGCTGTGCGCGCCCATTCGCACTGCTCCGAGAGAGAACATGC	1090
D	235245	GGCCCCAAGAACGAGCCAAAGCTGTGCGCGCCCATTCGCACTGCTCCGAGAGAGAACATGC	235186
OY	1091	GGGCG -TCCCTTCCGCCGGAAGGCGCACTGGGGCTCGCGGCTACACAACTGCTTGTATATT	1149
D	235185	GGGCTTCCCAACCCCGCAAGGTCACGTGGGGTTGCGCGCTTACAAACTGCTTGTATATT	235126
OY	1150	CTCGTTGGCCCTCTCACTCTGGCTTCCCGGCTACGCTATAGACAGTACCAGTTGTCT	1209
D	235125	CTCGTTGCTCTGACGCTGTGGCTTCCCGCTACGCTATGACAGTACCAGTTGTGCT	235066
OY	1210	TTTGCAAGTACTGGATCCCTTGGTCAAGCGGCTTTTCAAGGCGACAGCAGGATTAACG	1269
D	235065	TTTGGGAGCTACTGTGACCTTTTGTGTCAAGAGGCTTTTCAAGGCTACAGTGAAGCCAAAG	235006
OY	1270	TTTATGTATACAGAAATAGCAGACATCGCGCTGCTTATAGTATATAGTGGAGAGATGC	1329
D	235005	TTTATGTATACAGAAATAGTGGCGCATCTGCGCTGCTATGTGGTGTGTGGGAAATATGC	234946
OY	1330	AGGACCCGCTGGTGTGCGGCTGTGAGCGTGGGAAAGCAGTTGTATTCCTCGCACAT	1389
D	234945	AGGACCCGCTGTGTGCGGCTGTGAGCGTGGGAAAGCAGTGTATTCCTGTGCACACT	234886
OY	1390	GGCGGACGATGACACAAACATGTATCATCAATCTGTCACTAGCTAGATACAGAGA	1449
D	234885	GGAGGACGATGGGACAAACAGCTATTAACAACGTGTCCGGAAATGCAGACACAGAGA	234826
OY	1450	ACCTTCTCTATTAAGTCAAGTACTGGCGGTGCATAGTGGCGGCACTTCTTCACTG	1509
D	234825	ATCTACTCTTAAAGCTCAAGTCAAGCGGGGCACTAGTGGCGGCACTTCTTCACTG	234766
OY	1510	TTCAGTACAGACCTGGCTTGTACTTGGTGTGATATACCGCTGTGTCATGATGCTGAGC	1569
D	234765	CCCAATTAAGAGTGTGCTTTTACCTGTGTGTGTACACCCCTTTGTCAATGCTATGTGTAC	234706
OY	1570	CCAACTTATGAGAAATCCACCAAGGTGGCGGTGAAGCGGAAATATCTTTCACCTTCC	1629

Db 234705 CCAACTTCATGGAATCCCAACCGAGGTGCAGTTAAGCGGAAATATCTTTCACCTTCC 234646
 QY 1650 AGGGCGAGAAATATTGAGTCTCTTGAGGCTCTAGCTTTCAGGAGAGGCCGCTCTTGGAGAGG 1689
 Db 234645 AGGGCGAAGAAATGATGAGTCTCTGTGAGATCTAGCTTTCAGGAGAGGCCGCTCTTTCAGGAGAG 234586
 QY 1650 AAATGAGAGGGGACCTCTCCGCCGACATACATGATGACGGATCATTTGCCACCCCTGAAGGCG 1749
 Db 234585 AGATGAGAGGGGACCTCTCCGCCGACATGATGACGATGCCATTCATTGGCCACCCCTAAAGGCG 234526
 QY 1750 TGCAGACAGCAGCTGGATCAGGTCTCGGTGGAAATTCACCTG 1792
 Db 234525 TACAGACAGCAGCTGGATCAGGTCTCGGTGGAAATTCACCTG 234483

LOCUS	AC011641	64626 bp	DNA	linear	HTG 29-DEC-2000
RESULT 15	AC011641/c				
DEFINITION	Homo sapiens clone RP11-13N16, LOW-PASS SEQUENCE SAMPLING.				
ACCESSION	AC011641				
VERSION	AC011641.2 GI:11995638				
KEYWORDS	HTG; HTGS_PHASE0.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 64626)				
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.				
JOURNAL	Homo sapiens, clone RP11-13N16				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 64626)				
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,				
	Baldwin,J., Barna,N., Beckerley,R., Boguslavsky,L., Boukhalter,B.,				
	Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,				
	Cooke,P., Dearellano,K., Dewar,K., Domino,M., Doneelan,L., Doyle,M.,				
	Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,				
	Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,				
	Howland,J.C., Johnson,R., Jones,C., Kann,L., Karktas,A., Klein,J.,				
	Lehoczky,J., Lien,C., Locke,K., Macdonald,P., Marquis,N.,				
	McEwan,P., McGuck,A., McKernan,K., McLaughlin,J., Meldrum,J.,				
	Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,				
	Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Sevely,P.,				
	Stange-Thoman,N., Stojanovic,N., Subramanian,A., Talamas,J.,				
	Teifage,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,				
	Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.				
	Direct Submission				
COMMENT	Submitted (08-Oct-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA On Dec 29, 2000 this sequence version replaced gi:6016773. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html				
TITLE	Genome Center				
JOURNAL	Center: Whitehead Institute/ MIT Center for Genome Research				
	Center code: WIBR				
	Web site: http://www-seq.wi.mit.edu				
	Contact: sequence.submissions@genome.wi.mit.edu				
	Project Information				
	Center project name: L3369				
	Center clone name: 13_N16				
	NOTE: This record contains 78 individual				
	* sequencing reads that have not been assembled into				
	* contigs. Runs of N are used to separate the reads				
	* and the order in which they appear is completely				
	* arbitrary. Low-pass sequence sampling is useful for				
	* identifying clones that may be gene-rich and allows				
	* overlap relationships among clones to be deduced.				
	* However, it should not be assumed that this clone				
	* will be sequenced to completion. In the event that				
	* the record is updated, the accession number will				
	* be preserved.				
	1 734: contig of 734 bp in length				

```
*      735 834: gap of 100 bp
*      835 1573: contig of 739 bp in length
*      1574 1673: gap of 100 bp
*      1674 2401: contig of 728 bp in length
*      2402 2501: gap of 100 bp
*      2502 3211: contig of 710 bp in length
*      3212 3311: gap of 100 bp
*      3312 4048: contig of 737 bp in length
*      4049 4148: gap of 100 bp
*      4149 4850: contig of 702 bp in length
*      4851 4950: gap of 100 bp
*      4951 5675: contig of 725 bp in length
*      5676 5775: gap of 100 bp
*      5776 6464: contig of 689 bp in length
*      6465 6564: gap of 100 bp
*      6565 7266: contig of 702 bp in length
*      7267 7366: gap of 100 bp
*      7367 8098: contig of 732 bp in length
*      8099 8198: gap of 100 bp
*      8199 8908: contig of 710 bp in length
*      8909 9008: gap of 100 bp
*      9009 9733: contig of 725 bp in length
*      9734 9833: gap of 100 bp
*      9834 10557: contig of 724 bp in length
*      10558 10657: gap of 100 bp
*      10658 11380: contig of 723 bp in length
*      11381 11480: gap of 100 bp
*      11481 12202: contig of 722 bp in length
*      12203 12302: gap of 100 bp
*      12303 13058: contig of 756 bp in length
*      13059 13158: gap of 100 bp
*      13159 13902: contig of 744 bp in length
*      13903 14002: gap of 100 bp
*      14003 14723: contig of 721 bp in length
*      14724 14823: gap of 100 bp
*      14824 15543: contig of 726 bp in length
*      15550 15649: gap of 100 bp
*      15650 16389: contig of 740 bp in length
*      16390 16489: gap of 100 bp
*      16490 17192: contig of 703 bp in length
*      17193 17292: gap of 100 bp
*      17293 18006: contig of 714 bp in length
*      18007 18106: gap of 100 bp
*      18107 18812: contig of 706 bp in length
*      18813 18912: gap of 100 bp
*      18913 19655: contig of 743 bp in length
*      19656 19755: gap of 100 bp
*      19756 20490: contig of 735 bp in length
*      20491 20590: gap of 100 bp
*      20591 21335: contig of 745 bp in length
*      21336 21435: gap of 100 bp
*      21436 22140: contig of 705 bp in length
*      22141 22240: gap of 100 bp
*      22241 22963: contig of 723 bp in length
*      22964 23063: gap of 100 bp
*      23064 23780: contig of 717 bp in length
*      23781 23880: gap of 100 bp
*      23881 24616: contig of 736 bp in length
*      24617 24716: gap of 100 bp
*      24717 25484: contig of 768 bp in length
*      25485 25584: gap of 100 bp
*      25585 26329: contig of 745 bp in length
*      26330 26429: gap of 100 bp
*      26430 27188: contig of 759 bp in length
*      27189 27288: gap of 100 bp
*      27289 28037: contig of 749 bp in length
*      28038 28137: gap of 100 bp
*      28138 28886: contig of 749 bp in length
*      28887 28986: gap of 100 bp
*      28987 29720: contig of 734 bp in length
*      29721 29820: gap of 100 bp
*      29821 30566: contig of 746 bp in length
*      30567 30666: gap of 100 bp

*      30667 31409: contig of 743 bp in length
*      31410 31509: gap of 100 bp
*      31510 32229: contig of 720 bp in length
*      32230 32329: gap of 100 bp
*      32330 33057: contig of 728 bp in length
*      33058 33157: gap of 100 bp
*      33158 33856: contig of 699 bp in length
*      33857 33956: gap of 100 bp
*      33957 34683: contig of 727 bp in length
*      34684 34783: gap of 100 bp
*      34784 35513: contig of 730 bp in length
*      35514 35613: gap of 100 bp
*      35614 36344: contig of 731 bp in length
*      36345 36444: gap of 100 bp
*      36445 37202: contig of 758 bp in length
*      37203 37302: gap of 100 bp
*      37303 38042: contig of 740 bp in length
*      38043 38142: gap of 100 bp
*      38143 38869: contig of 727 bp in length
*      38870 38969: gap of 100 bp
*      38970 39724: contig of 755 bp in length
*      39725 39824: gap of 100 bp
*      39825 40595: contig of 771 bp in length
*      40596 40695: gap of 100 bp
*      40696 41435: contig of 740 bp in length
*      41436 41535: gap of 100 bp
*      41536 42234: contig of 699 bp in length
*      42235 42334: gap of 100 bp
*      42335 43069: contig of 735 bp in length
*      43070 43169: gap of 100 bp
*      43170 43878: contig of 709 bp in length
*      43879 43978: gap of 100 bp
*      43979 44708: contig of 730 bp in length
*      44709 44808: gap of 100 bp
*      44809 45526: contig of 718 bp in length
*      45527 45626: gap of 100 bp
*      45627 46349: contig of 723 bp in length
*      46350 46449: gap of 100 bp
*      46450 47182: contig of 733 bp in length
*      47183 47282: gap of 100 bp
*      47283 48007: contig of 725 bp in length
*      48008 48107: gap of 100 bp
*      48108 48827: contig of 720 bp in length
*      48828 48927: gap of 100 bp
*      48928 49651: contig of 724 bp in length
*      49652 49751: gap of 100 bp
*      49752 50470: contig of 719 bp in length
*      50471 50570: gap of 100 bp
*      50571 51314: contig of 744 bp in length
*      51315 51414: gap of 100 bp
*      51415 52148: contig of 734 bp in length
*      52149 52248: gap of 100 bp
*      52249 52982: contig of 734 bp in length
*      52983 53082: gap of 100 bp
*      53083 53779: contig of 697 bp in length
*      53780 53879: gap of 100 bp
*      53880 54593: contig of 714 bp in length
*      54594 54693: gap of 100 bp
*      54694 55420: contig of 727 bp in length
*      55421 55520: gap of 100 bp
*      55521 56249: contig of 729 bp in length
*      56250 56349: gap of 100 bp
*      56350 57095: contig of 746 bp in length
*      57096 57195: gap of 100 bp
*      57196 57942: contig of 747 bp in length
*      57943 58042: gap of 100 bp
*      58043 58771: contig of 729 bp in length
*      58772 58871: gap of 100 bp
*      58872 59586: contig of 715 bp in length
*      59587 59686: gap of 100 bp
*      59687 60408: contig of 722 bp in length

10.1%; Score 621; DB 2; Length 64626;
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 4, 2002, 14:58:36 ; Search time 15.05 Seconds

(without alignments)
2364.337 Million cell updates/sec

Title: US-09-809-920-4

Perfect score: 4873
Sequence: 1 MTGYTMRNGAGNGGQTCM.....DSVLEKTRLPDHDKCEKFI 919

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4873	100.0	919	1	EXL3_HUMAN
2	1251.5	25.7	814	1	EXT2_MOUSE
3	1004	20.6	718	1	EXT2_MOUSE
4	1000	20.5	718	1	EXT2_MOUSE
5	871.5	17.9	746	1	EXT1_MOUSE
6	871	17.9	746	1	EXT1_MOUSE
7	660.5	13.6	676	1	EXL1_HUMAN
8	367	7.5	330	1	EXL1_HUMAN
9	251.5	5.2	378	1	EXT1_MOUSE
10	123.5	2.5	622	1	VP27_YEAST
11	113	2.3	676	1	VATI_ARCFU
12	113	2.3	2663	1	CENE_HUMAN
13	112.5	2.3	4568	1	DYHC_CAEEL
14	111.5	2.3	3075	1	LMAL_HUMAN
15	110.5	2.3	429	1	CPXR_BRAJA
16	110.5	2.3	1940	1	MYH3_CHICK
17	109	2.2	978	1	RA50_AQUAE
18	109	2.2	4967	1	RYR2_HUMAN
19	108.5	2.2	1311	1	FMR2_HUMAN
20	108.5	2.2	1595	1	SOS_DROME
21	108	2.2	873	1	PC1_HUMAN
22	107.5	2.2	1087	1	OAS3_HUMAN
23	107	2.2	1748	1	POLR_ELV
24	107	2.2	4684	1	PLEI_HUMAN
25	106.5	2.2	700	1	HS9C_DICDI
26	106.5	2.2	2324	1	COAC_CHICK
27	106	2.2	732	1	ATZN_ECOLI
28	105.5	2.2	1976	1	MYHA_BOVIN
29	105.5	2.2	2871	1	DESP_HUMAN
30	105	2.2	474	1	YLEA_HAEIN
31	104.5	2.1	1505	1	CUT2_HUMAN
32	104.5	2.1	4128	1	PRKD_HUMAN
33	104	2.1	359	1	PEXC_CAEEL

ALIGNMENTS

RESULT	1	STANDARD	PRT	919 AA.	
EXL3_HUMAN					
ID	EXL3_HUMAN	043909: 000225:			
AC	15-JUL-1999 (Rel. 38, Created)				
DT	15-JUL-1999 (Rel. 38, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Exostosin-like 3 (putative tumour suppressor protein EXTL3) (multiple exostosin-like protein 3) (Hereditary multiple exostososes gene Isolog)				
DE	(EXT-related protein 1)				
GN	EXTL3 OR EXTRL OR EXTLR OR KIAA0519.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
OX	NCBI TaxID: 9606;				
RA	11				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-98140122; Pubmed-9479495;				
RA	Ivan Hui W., Wuys W., Hendrickx J., Speleman F., Wauters J.,				
RA	de Boule K., van Roy N., Bossuyt P., Williams P.J.;				
RT	"Identification of a third EXT-like gene (EXTL3) belonging to the EXT				
RT	gene family.";				
RL	Genomics 47:230-237(1998).				
RN	121				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-Testis, and Fetal;				
RX	MEDLINE-98139867; Pubmed-9473480;				
RA	Sato T., Seki N., Yamauchi M., Tsuji S., Hayashi A., Kozuma S.,				
RA	Hori T. A.;				
RT	"Structure, chromosomal location, and expression profile of EXTL1 and				
RT	EXTL2, new members of the multiple exostososes gene family.";				
RL	Biochem. Biophys. Res. Commun. 243:61-66(1998).				
RN	131				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-Brain;				
RX	MEDLINE-98290545; Pubmed-9628581;				
RA	Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,				
RA	Nomura N., Ohara O.;				
RT	"Prediction of the coding sequences of unidentified human genes. IX.				
RT	The complete sequences of 100 new cDNA clones from brain which can				
RL	code for large proteins"-in vitro.";				
RL	DNA Res. 5:31-39(1998)				
RN	141				
RP	SEQUENCE FROM N.A.				
RA	Deng H.-X., Xu L., Xia J.-H., Fan C., Pan Q., Liu C.-Y., Ruan Q.-G.;				
RT	"Molecular cloning of a candidate gene for multiple exostososes.";				
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.				
RN	151				
RP	SEQUENCE FROM N.A.				
RA	Sato T.;				
RT	"The human EXTL1/EXTL2/EXTL3 gene at 8p11-p12, the third breast				
RT	cancer susceptibility gene locus, is not mutated in breast and various				
RL	cancers.";				
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.				
RN	161				
RP	SEQUENCE OF 470-919 FROM N.A.				

RA Adams M.D.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PROBABLE TUMOR SUPPRESSOR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
 CC reticulum (By similarity).
 CC -1- TISSUE SPECIFICITY: FOUND IN ALL TISSUES TESTED.
 CC -1- DISEASE: CANDIDATE GENE FOR THE BREAST CANCER LOCUS ON CHROMOSOME
 CC 8p12-p22.
 CC -1- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
 CC CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-6 OR MET-20 IS THE
 CC INITIATOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF001690; AAC39598.1; -
 DR EMBL; AF007042; BAA24080.1; -
 DR EMBL; AB011091; BAA25445.1; -
 DR EMBL; AF029231; AAD01877.1; -
 DR EMBL; AF083531; AAD42041.1; -
 DR EMBL; U96629; AAB67602.1; ALT_SEQ.
 DR MIM; 605744; -
 DR InterPro: IPR004263; Exostosin.
 DR Pfam: PF03016; Exostosin; 1.
 DR Anti-oncogene; Multigene family; Transmembrane; Signal-anchor.
 KW TRANSFER 31
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 692 695
 FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 919 AA; 104748 MW; 200AD4DAB4A39FD CRC64;

Query Match 100.0%; Score 4873; DB 1; Length 919;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 919; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGTYMLRNGAGNGGOTCMRLMSNRIRLWLSFTLVIIIVFFPLIAHYLTDEDEA 60
 DB 1 MTGTYMLRNGAGNGGOTCMRLMSNRIRLWLSFTLVIIIVFFPLIAHYLTDEDEA 60
 Y 61 GKRIFFRGVNEELCEVKNVLDLCIRISVSBEILLQLEAKROELNSETAKTNLKEACKS 120
 DB 61 GKRIFFRGVNEELCEVKNVLDLCIRISVSBEILLQLEAKROELNSETAKTNLKEACKS 120
 Y 121 IENAKODLLQKNTISTEHSYKELMNONPKLSLPIRLPEKDDAGLPKPKATRGCRH 180
 DB 121 IENAKODLLQKNTISTEHSYKELMNONPKLSLPIRLPEKDDAGLPKPKATRGCRH 180
 Y 181 NCFDYSNCPILSGFPVYVSDQFVFGSYLDPLVKAQFAQATARANYVTENADIACYVI 240
 DB 181 NCFDYSNCPILSGFPVYVSDQFVFGSYLDPLVKAQFAQATARANYVTENADIACYVI 240
 Y 241 LVGEMQEPVYLRAPELEKQYSLPHMFTDGHNNHYINLSRKSPTQNLNLYVSTGRANVAQ 300
 DB 241 LVGEMQEPVYLRAPELEKQYSLPHMFTDGHNNHYINLSRKSPTQNLNLYVSTGRANVAQ 300
 Y 301 STFTYVYRGGFDLVSPVLANHSEPFMEIPPOVPYKRYLFTFOGEKTESLRSSIQEA 360
 DB 301 STFTYVYRGGFDLVSPVLANHSEPFMEIPPOVPYKRYLFTFOGEKTESLRSSIQEA 360
 Y 361 RSFEENEKGPDPADYDRITATLKAODSKLDVLFETCKNPKPSLPTFEMALGGERD 420
 DB 361 RSFEENEKGPDPADYDRITATLKAODSKLDVLFETCKNPKPSLPTFEMALGGERD 420
 Y 421 RLELLKLTSTFALITTPGDPRLVYSSGCATRLFEALEVGAVPVVLGEOVQDLPYQDMLQNE 480

DB 421 RLELLKLTSTFALITTPGDPRLVYSSGCATRLFEALEVGAVPVVLGEOVQDLPYQDMLQNE 480
 Y 481 AALVPRPRVEVHFLRLSUSDLLAMRRGCRFLMEYESTASINTVYAMIRTRIQI 540
 DB 481 AALVPRPRVEVHFLRLSUSDLLAMRRGCRFLMEYESTASINTVYAMIRTRIQI 540
 Y 541 PAAPIREBAAEIHRSGKAGTDPNNMADNGDLGFEVETPEPPASPRYLRNFTLVYTD 600
 DB 541 PAAPIREBAAEIHRSGKAGTDPNNMADNGDLGFEVETPEPPASPRYLRNFTLVYTD 600
 Y 601 YRSNCAAGPFFHLPFPDPVLPSEAKFLGSGTGFPPITGGAGSGKEQALGAVPR 660
 DB 601 YRSNCAAGPFFHLPFPDPVLPSEAKFLGSGTGFPPITGGAGSGKEQALGAVPR 660
 Y 661 EGFVVMVLTREBEVLNLSERLGLYLKVVVWVNSPKLPSSDLMPDGVITVVRT 720
 DB 661 EGFVVMVLTREBEVLNLSERLGLYLKVVVWVNSPKLPSSDLMPDGVITVVRT 720
 Y 721 EKNSLNRFLPWNIEETALISIDDAHLRDEIMFGFRVREARDRIVGPGRYHAMDI 780
 DB 721 EKNSLNRFLPWNIEETALISIDDAHLRDEIMFGFRVREARDRIVGPGRYHAMDI 780
 Y 781 PHOSMLYNSNSCELSWLTGAFFHKYAYLYSVMPQAIRDMVDEYINCEDIAMNELY 840
 DB 781 PHOSMLYNSNSCELSWLTGAFFHKYAYLYSVMPQAIRDMVDEYINCEDIAMNELY 840
 Y 841 SHTRKRPPIVTSWMTFRGCGQALSHDSSHFERHKCINFPKYVGYMPLLYTORVD 900
 DB 841 SHTRKRPPIVTSWMTFRGCGQALSHDSSHFERHKCINFPKYVGYMPLLYTORVD 900
 Y 901 SVLEKTRLPDHTKCFKFI 919
 DB 901 SVLEKTRLPDHTKCFKFI 919

RESULT 2
 EXT2 CAEEL STANDARD; PRT; 814 AA.
 AC 001705;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE Multiple exostosins homolog 2.
 GN RIB-2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodera; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN RP
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9726342; PubMed=9110175;
 RA Clines G.A., Ashley J.A., Shah S., Lovett M.;
 RT "The structure of the human multiple exostosins 2 gene and
 RT characterization of homologs in mouse and Caenorhabditis elegans.";
 RL Genome Res. 7:359-367(1997).
 CC -1- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U94835; AAC47510.1; -
 DR InterPro: IPR004263; Exostosin.
 DR Pfam: PF03016; Exostosin; 1.
 SQ SEQUENCE 814 AA; 94182 MW; 0A9A2AC7CE308E14 CRC64;

Query Match 25.7%; Score 1251.5; DB 1; Length 814;

Best Local Similarity 33.4%; Pred. No. 3 6e-82;
Matches 313; Conservative 147; Mismatches 312; Indels 165; Gaps 25;

```

OY 21 LRMSNRIRLTWLSFTFLVLPPEL--IAHYLT-----TDEADEACKRI 64
DB 5 LNSSRSFVPSLRSAFLIFFFIYIILYNVSFSEPSMITODALKONENLDDYASC 64
OY 65 FGPVNGELCEVKNVLDLCRIRESVSE-ELLQLEAKROELNSELAKNLKIEACKSIEN 123
DB 65 SGYSIGRILREOKRILASVRELSESVKIEERTVOEQLRLLPOKOLELSALEGIEIA 124
OY 124 AKODLLQKLVISQTESHYKELMAQNPRLSLRLLPEKDDAGLPPKATRCGRINCFC 183
DB 125 AQORL-----BELRQTONVKVFLPRS--PLQIPRELEQSPQISPNQDLII 168
OY 184 DYSRCLTSGFPVYVYDSDQFVFGSYLDPLVKAQFQATARANYVTENADIACLYILVG 243
DB 169 DYSRCSISSEFPVYV-----FOSKSIHQASKI--RSF 259
OY 244 EMQEPVYLRPALEKOLYSLPHMRTDGHNVIIILSKSDTONLLVNSVGRANVAOSTF 303
DB 184 ----DIITSGQSEKELNV-----FOEYIPNLVETPKACIKIRISNGIAS--PMTTF 230
OY 304 YTVQYRGFDLVVSPLVHAHSEPNFMEIPQVPYKRYLFTFOGKIESLSQEARSF 363
DB 231 NSILFNNG-----SPIIN-----FOSKSIHQASKI--RSF 259
OY 364 EEEMEGDPPADYD---DRI-IATLKAQODSKLDQVLVEFTCKN-----QPKPS-LPT 410
DB 260 -----DPPVDVNHIAVEKVDLPFLPFORENLSILVDNTELFSAFSSLSAPSRRI 313
OY 411 EWALCGR---EDRLKLTSTFALITTPEDPLVYSSGCANRLFEALEGAVPVVLGE 466
DB 314 VIVKCSQENCSLERRDILSSSTFCFLP-----SEMFQDFLSLQGCIPILSN 365
OY 467 QVQVLPYQDMQWNAALVVKPRVTEVHFLLRSLSDDLAMROGRFMEITYSTADSI 526
DB 366 SOLPLPFDLDMRRATYRLPLARLEAHFTVQSEISDILIMRVGGLFETYLAHRHL 425
OY 527 FNTVLAIRTRIQIPAPIREEAAEIPRSGKA-AGTDPMADNGDLD--LGFVETEP 583
DB 426 ARSLALARKKLOJPTREVRNQAIPFNSSFTAPKGSVNVQANPDEYVLGLSESR-- 483
OY 584 YASGRYLRNFTLVTDYFYSKNCACGPHLPPHTPDPVLPSEAKL-GSGTGERPIGG 642
DB 484 VESTSYAVNTEFQVLSYDPMNIIIMSPHYTKFELVNAALPTEAEFPEDKIGFRPI--- 540
OY 643 AGSGRKEFOALGQNVNREQFTVYMLTYERREVLNMLERLNGLPYLNKVVYVANSKLP 702
DB 541 EPGSGAEFSKALGQNRQREQFTVYLLTYERDAVLTGALERHLOLPYLNKIIVANNVNRD 600
OY 703 SEDLWPDIGVPIVWYRTEKSLNRFPLMNEIETEAIIISIDDAHLRHEIMFGFVWR 762
DB 601 PPD-SWESLIHPVFIVAEENLNKRPVMDRITEAVLSIDDDIDIMODEIILAFVWR 659
OY 763 EARDRIVGFGYRHAWDIPHOSWLYNSYSCSELSVLTGAFFHKYAYLYSVYVQAIR 822
DB 660 ENRDRIVGFPARHHA--RYGDSMEFYNSNHTCOMSMILTGAFFHKYAYLYVEMPAIR 717
OY 823 DMVDEYINCEDIAMNELYSHTRKPRPKYISRWTFRCGPGQALSHDSHFHEHAKCIN 882
DB 718 EHVNSINCEDIAMNELYSHLTRKPRPKYISRWTLKCPCTESLYKGTGHEKHECMRL 777
OY 883 FVKVGYMPLLYTOFRVDSVLFKTRLPDHTKCFKFI 919
DB 778 FTKYIGYNPLKFSQFRADSLFKTRLPQNHQCKRYV 814

```

RESULT 3
EXT2_MOUSE STANDARD; PRT; 718 AA.
AC P70428: P70395;
DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Exostosin-2 (Putative tumor suppressor protein EXT2) (Multiple
exostoses protein 2).
GN EXT2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97376020; PubMed=9232192;
RA Stickens D.J., Evans G.A.,
RT "Isolation and characterization of the murine homolog of the human
EXT2 multiple exostoses gene."
RL Biochem. Mol. Med. 61:16-21(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97264342; PubMed=9110175;
RA Clines G.A., Ashley J.A., Shah S., Lovett M.,
RT "The structure of the human multiple exostoses 2 gene and
characterization of homologs in mouse and Caenorhabditis elegans."
RL Genome Res. 7:359-367(1997).
CC -1- FUNCTION: APPEARS TO BE A TUMOR SUPPRESSOR.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
reticulum (By similarity).
CC -1- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: U72141; AAB17006.1; -;
CC EMBL: U67837; AAC5143.1; -;
CC DR MGD: MGI:108050; Ext2.
CC DR InterPro: IPR004263; Exostosin.
CC DR Pfam: PF03016; Exostosin; 1.
CC KW Anti-oncogene; Multigene family; Transmembrane; Signal-anchor
TRANSMEM 26 46
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 637 637 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 62 62 S -> T (IN REF. 2).
FT CONFLICT 160 160 L -> V (IN REF. 2).
FT CONFLICT 355 355 V -> I (IN REF. 2).
FT CONFLICT 389 389 Q -> H (IN REF. 2).
SQ SEQUENCE 718 AA; 82105 MW; C2E765393A7F7ABD CRC64;

Query Match 20.6%; Score 1004; DB 1; Length 718;
Best Local Similarity 32.2%; Pred. No. 1 9e-64;
Matches 260; Conservative 126; Mismatches 254; Indels 168; Gaps 27;

```

OY 133 NVISQTESHYKELMAQNPRLSLRLLPEKDDAGLPPKATRCGRHNCFPVSRCLPT 191
DB 49 SISSSDGVEKKSIRKVPYVRLPTDSPIPERGD-----SCMHATCFDYKRC--- 96
OY 192 SGR-----PVVYDSDQFV--FGSYLDPLVKAQFQATARA---NVVTENADIACLYI 240
DB 97 -GPNPKRIKIVYIYPLKKYVDGAGVPYSAISREYNELLTAISDSYDITDIRACLFVP 155
OY 241 LVGEM-QEPVYLRPALEKOLYSLPHMRTDGHNVIIIN-LSKSDTONLLVNSVGRANV 298
DB 156 SIDVLNQP--LRKETAQALQSLRW-DRGTNHLFLNMLPGAPPDYNTALVPRDRAAL 212
OY 299 AOSTFYVQYRPGFDL---VSPVLVHAHSEPNFMEIPQVPYKRYLFTFOGKIESLSRS 355
DB 213 AGGFGSTWYTRQGYDSIPVFSLSAEMA-----LPEKAPGRRIFL-----LSS 257

```

```

QY 356 SLQARSFEEMEGDPPADYDRIATATKAQDSKLDQVYVEFTCKNPKPSLPTMALC 415
Db 258 QMAHPEYRELE-----ALQAKHQSIVLADKCTNSEGLSVR-KRC 300
QY 416 GERE--DRLELLKLTSTFALLITPGDPLVSSGATRLFELEVGAVVVGEOQLRYQ 473
Db 301 HOHGVDFDPOVLAQENTFCTVLRRA--RL-----GQAV-LSDVLAQACVAVVADSTILPFS 353
QY 474 DMLQNEALVYKPRVTEVHFLRLSISDSLLAMRQGRFLMETYSTADSIENFTVLAM 533
Db 354 EVLDMKASVYVPEEKMDVYSILQNIPOQRIEEMQROARFWEAYFQSTAKALATLQI 413
QY 534 IRTIQTIPAPIRREAAAEIHRSGKAGTIPNMADNDLGLVETEPPTASPRYLNK 593
Db 414 INDRY-----PYAAIS----- 425
QY 594 TLFTVDFRSWNCAPGPHLPHP-PPVLPSEAKFLGSGTGRPGIGAGSGSKERQA 652
b 426 -----YEENDDPPAKVMAVSNDPLPLIPPOS----- 453
QY 653 ALGNVPREQFTVYMLTYEREVEVLSNLERLNGLPYLKVVVWNSP-KLPSDDLMPDI 711
Db 454 -----OGFTALVLTFRVESLFRVLTVEYSKPSLSKLLVWNNQKNPESIMPKI 505
QY 712 GVPIMVVRTEKNSLNNRFLPNNEIETETAILSIDD-AHLRDEIMFGFRVREARDIVG 770
Db 506 RVPKRVTAETAKLSNRFPPDEIETETAVLAIDDIIMLTSDLOFGYEVAREFPDRLVG 565
QY 771 FPGRYHAMDIPOHSMLYNSYSCSLMWLTGAFFHKYAYLYSVMPQALRDWVDEYN 830
Db 566 YPGSLHMDHDMKNKKESEWTEVNSVAVLGAALYHKYFNLYYKMGDGIKNVADAHN 625
QY 831 CEDIANNFVSHITRKPPIKVTSWTFRCGCP--QALSHDSHFHEKHCINEFVYVG 888
Db 626 CEDIANNFVANYGKAVIKVTPRKKECECTAIDGLSDQTHVRSSECKINFASVFG 685
QY 889 YMPLYTQFRVSYLFRTRLPHDKTGF 916
Db 686 TMLKVEYHRADPVLTKDDP-EKLKSF 712

```

RESULT 4
EXT2_HUMAN
ID EXT2_HUMAN STANDARD; PRT; 718 AA.
AC Q93063;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Exostosin-2 (Putative tumor suppressor protein EXT2) (Multiple
exostosins protein 2).
GN EXT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96376967; PubMed=8782816;
RA Stikens D.J., Clines G., Burbee D.G., Ramos P., Thomas S., Hogue D.,
RA Hecht J.T., Lovett M., Evans G.A.;
RT "The EXT2 multiple exostosins gene defines a family of putative tumour
suppressor genes.";
RT Nat. Genet. 14:25-32(1996).
RL [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=97049960; PubMed=8894688;
RA Buys W., van Hul W., Wauters J., Nemtsova M., Reyniers E.,
RA van Hul E., de Bouille K., de Vries B.B.A., Hendrickx J., Herrygers I.,
RA Bossuyt P., Bailemans W., Franssen E., Vits L., Coucke P., Nowak N.J.,
RA Mallet L., van den Ouweland A.M.W., McGaughan J., Halley D.J.J.,
RA Willems P.J.;
RT "Positional cloning of a gene involved in hereditary multiple

```

RT exostosins.";  
RL Hum. Mol. Genet. 5:1547-1557(1996).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97264342; PubMed=9110175;  
RA Clines G.A., Ashley J.A., Shan S., Lovett M.;  
RT "The structure of the human multiple exostosins 2 gene and  
characterization of homologs in mouse and Caenorhabditis elegans.";  
RL Genome Res. 7:359-367(1997).  
RN [4]  
RP VARIANT EXT2 ASN-227.  
RX MEDLINE=97465597; PubMed=9326317;  
RA Philippe C., Porter D.E., Emerton M.E., Wells D.E., Simpson A.H.R.W.,  
RA Monaco A.P.;  
RT "Mutation screening of the EXT1 and EXT2 genes in patients with  
hereditary multiple exostosins.";  
RL Am. J. Hum. Genet. 61:520-528(1997).  
CC -1- FUNCTION: APPEARS TO BE A TUMOR SUPPRESSOR.  
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic  
CC reticulum (By similarity).  
CC -1- DISEASE: DEFECTS IN EXT2 ARE ASSOCIATED WITH HEREDITARY MULTIPLE  
CC EXOSTOSIS (EXT). EXT IS A GENETICALLY HETEROGENEOUS BONE DISORDER  
CC CAUSED BY GENES SEGREGATING ON HUMAN CHROMOSOMES 8, 11, AND 19 AND  
CC DESIGNATED EXT1, EXT2 AND EXT3 RESPECTIVELY. EXT IS A DOMINANTLY  
CC INHERITED SKELETAL DISORDER PRIMARILY AFFECTING ENCHONDRAL BONE  
CC DURING GROWTH. THE DISEASE IS CHARACTERIZED BY FORMATION OF  
CC NUMEROUS CARTILAGE-CAPPED, BENIGN BONE TUMORS THAT ARE OFTEN  
CC ACCOMPANIED BY SKELETAL DEFORMITIES AND SHORT STATURE. IN A SMALL  
CC PERCENTAGE OF CASES EXOSTOSIS HAVE EXHIBITED MALIGNANT  
CC TRANSFORMATION RESULTING IN AN OSTEOSARCOMA OR CHONDROSARCOMA.  
CC -1- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC  
DR EMBL: U62740; AAB07008.1; -;  
DR EMBL: U64511; AAC50764.1; -;  
DR EMBL: U67368; AAC51219.1; -;  
DR EMBL: U67356; AAC51219.1; JOINED.  
DR EMBL: U67357; AAC51219.1; JOINED.  
DR EMBL: U67358; AAC51219.1; JOINED.  
DR EMBL: U67360; AAC51219.1; JOINED.  
DR EMBL: U67361; AAC51219.1; JOINED.  
DR EMBL: U67362; AAC51219.1; JOINED.  
DR EMBL: U67363; AAC51219.1; JOINED.  
DR EMBL: U67364; AAC51219.1; JOINED.  
DR EMBL: U67365; AAC51219.1; JOINED.  
DR EMBL: U67366; AAC51219.1; JOINED.  
DR EMBL: U67367; AAC51219.1; JOINED.  
DR MIM: 133701.  
DR InterPro: IPR004263; Exostosin.  
DR Pfam: PF03016; Exostosin; 1.  
KW Anti-oncogene; Multigene family; Transmembrane; Signal-anchor;  
KW Disease mutation.  
FT TRANSMEM 26 46  
FT CARBOHYD 288 288  
FT CARBOHYD 637 637  
FT VARIANT 227 227  
SQ SEQUENCE 718 AA; 82254 MW; 9048CD3A5B63C5CB CRC64;  


```

Query Match 20.5%; Score 1000; DB 1; Length 718;
Best Local Similarity 32.5%; Pred. No. 3, 7e-64;
Matches 260; Conservative 125; Mismatches 258; Indels 156; Gaps 27;
QY 140 HSYKEIMAOHQKPLSL--PIRLPEKDAGLP-PPKATRGRLHNCFDYSRCPLTSGP- 194

```

Db 48 HSTESSNDMVEKRSTIDVYVRLP-----ADSIPIPERGDLSCNMHTCFDYRC-----GFN 99
OY 195 -----PVYVYDSQFV--FGSYLDPLVYKQAFQATARA---NVYVENADIACLVILVGE 244
Db 100 PKKIKYIYALKKYVDDEFSVSNITSREYNELMALISSDVTYTDINKACLFVPSIDV 159
OY 245 MOEPVYLRPELEKOLYSLPHMRTDGHNVITN-LSRKSPTONLLVNSGKRAMVAOSTF 303
Db 160 LNQN-TLRKETQAAMQALSRM-DRGNHLLFLNMLPGPPDYNTALDVPDRALLAGGCF 217
OY 304 YTVQYRPGDPLVSPLVHANSERNFMEIRPOVYVKKRYLFTPGGEKIESRSSLSQEARSF 363
Db 218 STYTVRQGVDSI-PVYSPLSAE--VDLPKGGPQYFV-----LSSQVGLHPEY 265
OY 364 EEEEMGGPRADYDRIATLAKAVODSKLDVLEFCTKNQPKPSLPTEMALCGERE--DR 421
Db 266 REDELE-----ALQVKGESVLYLDKCTNLSEGLSVR-KRCHKHOFDY 308
OY 422 LELLKSTFALLITPGDPLVYSSGATRLFEALVGVAVPVVLEGOVLPYDMLQWNEA 481
Db 309 PVLQGEATFCVLR--GARL---GGAV-LSQVLAQACVPIVADSYILPFESEVLDKRA 361
OY 482 ALVYPRPVTEVHFLRLSLSDLLMRGGRFLMETESTADSIPTVYLAIRTRIQIP 541
Db 362 SVYVPEKMSDVYSILQSIQROIEMQORARFWEAYFQSIKAIALATLQIINDRIYYP 421
OY 542 AARIREEAAEIPHRSGKAGTDPNMADNGDLDPVETPEPYASPRYLNNFTLVDFY 601
Db 422 ALSTYEE-----WNPRAV----- 435
OY 602 RSMNCAGRPFLHPHPEDVLPSEAKFLGSGFRPDIGGAGSGKEFOALGCVNPRE 661
Db 436 -KMGVSNPFLF-----PLIPPOG-----Q 454
OY 662 QFVVMVLTREVEVANSLERLNGLPYLNKVYVVMNSP-KLPSEDLMPDIOGPIVYVRT 720
Db 455 GFATVLTQYRVESLFRVITEVSKVPSLSKLLVVMNQNNPPEDSLMPKIRVPLKVVRT 514
OY 721 EKSLSNRLFLPMNIEETEAISIDDD-AHLRDEIMFGFVWREARIRYIGFPGRYHAM 779
Db 515 AEKLSRPFYDEIETEVLAIDDDIIMLTSDLOCGYEVWREFRPLVGYGRHLMD 574
OY 780 IPIQSWLYNSVSCELSMVLTGAAFEHKKYAYLYSVYVPOAIRDMVDEYINCEDIAMFL 839
Db 575 HEMNKMVEESEMTEVSMVLTGAAFYHKKYFNLYTYKMPGDIKMWDAHNNCEDIAMFL 634
OY 840 VSHITRPRPKYVSRWTFRCGCR--QALSHDSSHFERHKCTNFYKVVGYVMPLYTQF 897
Db 635 VAVVTGKAVLTVPKPKKFCPECTALDLSLDQTHMVERSECINKFASVEGTMLKVEH 694
OY 898 RVDSVLEKTRLPKHDTKCF 916
Db 695 RADPVLYLKKDFFP-EKLKSF 712

```

```

RX MEDLINE-97330052; PubMed-9186511;
RA Lohmann D.R., Bulding K., Lueddecke H.J., Horsthemke B.;
RT "The murine Ext1 gene shows a high level of sequence similarity with
RT its human homologue and is part of a conserved linkage group on
RL chromosome 15";
RN Cytogenet. Cell Genet. 76:164-166(1997).
RP SEQUENCE FROM N.A.
RC STRAIN-C57BL/6;
RX MEDLINE-97396026; PubMed-9254013;
RA Lin X., Wells D.;
RT "Isolation of the mouse cDNA homologous to the human EXT1 gene
RT responsible for hereditary multiple exostoses";
RL DNA Seq. 7:199-202(1997).
CC -1- FUNCTION: APPEARS TO BE A TUMOR SUPPRESSOR.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC reticulum (By similarity).
CC -1- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: U78539; AABA1728.1; -.
DR EMBL: X96639; CAA65443.1; -.
DR MGI: 894663; Ext1.
DR InterPro: IPR004263; Exostotin.
DR Pfam: PF03016; Exostotin; 1.
KW Anti-oncogene; Multigene family; Transmembrane; Signal-anchor.
FT TRANSMEM 8 28
FT FT
FT CARBOHYD 89 89
FT CARBOHYD 330 330
FT CONFLICT 16 16
FT CONFLICT 60 60
FT CONFLICT 220 220
FT CONFLICT 486 486
FT CONFLICT 548 548
SQ SEQUENCE 746 AA; 86307 MW; 5AC7F24BCEDEF9 CRC64;

Query Match 17.9%; Score 871.5; DB 1; Length 746;
Best Local Similarity 28.6%; Pred. No. 7e-55;
Matches 240; Conservative 142; Mismatches 245; Indels 211; Gaps 33;

OY 139 EHSYKELMAONQK---LSLPIRLLP-----EKDDAGLP-PRKATRG-----C 177
Db 39 EHSGRNGLHQPSRDHWRPRPDALRPFPPWDOLENEDSSVHSRKRDNSSYYKSKC 98
OY 178 RLHNCPRYSRCPRLSGFRVYV--DSQDFVFGSYLDPLVYKQAFQATARAUVYVENADIA 235
Db 99 RMESCFDEFTLCK-KNGKVVYVYPOQKGEKTAEST-----QNLIAEGSRFTTSDSQ 151
OY 236 CLVYIIVGEMQEPVYLRPA--ELEKOLYSLPHMRTDGHNVITN---SRKSDTONLLY 289
Db 152 CLFVLSIDTLDRD-QLSPOVYVHNRKSVQSLHLM-NNGRNHLLFLNLYSGWIPDTEVGF 209
OY 290 NVSTGRAMVAOSTFYVYVYRPGDPLVSPV---HAMSEPNMELIRPOVYVKKRYLFT 344
Db 210 DI--QAMLAKKASISTENFRNFDVSI-PLFSKDHPRGTGGERGLKP-NTIPRLRYKMLV 265
OY 345 FQGEKI-----ESLRSSIQARSEEMEDGPRADYDRIATLTK-----AVODSKLD 393
Db 266 FKGRKRYLTGSDTRNLVYVHNGEDVL-----LITTCCKHGKDMQKNDSCDR 314
OY 394 VIVEFTCKNQPKPSLPTEMALCGEREDRLKSTFALLITPGDPLVYSSGATRLPE 453
Db 315 DNEY-----EKDYREMLNHATGCLV--PRGRRL-----GSFRFLE 349

```

QY	454	ALEVGAVPVVLGEQVQLPYQDMLQNEALVYKPRVRYTEVFLKRLSDSDLLAMRQGR	513
Db	350	ALQAAACVPMVLSSNGWELPSEYVNNQAAGVDEKLLDLPSTRTSRISHODKILALROOTQ	409
QY	514	FLMETVYFSTASISFNTVLMATRTRIQIOPAADIREBAAAEIPIHRSKGAAGTDPNNAADGDL	573
Db	410	FLMEAVFSSVEKIVLTLEIITQDR-----FHHIS-----	439
QY	574	DIGVETEPBPYASPRYLNRFTLVYDTEFRSNNCAGPETHLPHTPPDPVLPSKAEFLG--	631
Db	440	-----RNSLI-----WNKHPG--GLF-----VLPOYSSVIGDF	465
QY	632	-----SGTGRPPIGGAGSGSGKEFOAALGCGNPRQGFVYMLITEREYVLANSLERLNGLP	667
Db	466	PRYANLCLKP-----PSKFAVLHAATP-----LVQSOPVYKLKLVAAAKSQ	508
QY	668	VLNKVVVVVWNSPK-LPSEDLMPDIGVIMVYRTKNSLNNRPLPMNEIFTEALISTDD	746
Db	509	YCAQITVLNMCKDPLPAHH-RWPAITAPVYIIEBSKYSKMSRFLPDNITTDVLSLDED	567
QY	747	AHLRHDELMGEFRVYRREARDRIVGPGERYHAMDIPHOQSWLTVNSYSCELSMVLGAAFFH	806
Db	568	TVLSTTEVDFAFTWQSPFERIVGYPARSHFMDNSKERMGVYTSKMTNDYSMLVLGAAYH	627
QY	807	KYVYVIVSYVWPAQAIRMDVETINCEDIANELVSHITRKRPKYV-----	852
Db	628	KYHVLVSHYIPALKNNVNDLACEDILNLFVSAVTKLPKLYTKVOKKOYKRETMAGQTS	687
QY	853	--SRMTFCPCGPOALSHDSSHFERHCKCINFPVKYVGYMPLLYTQFRVDSVLEFKRL	908
Db	688	RASMA-----DDPHFAQRQSCAMTASMGFTYPLTHSQMRDLPVLEFKOV	733
RESULT	6		
EXNT_HUMAN	STANDARD;	PRT;	746 AA.
ID	EXNT_HUMAN	STANDARD;	PRT;
AC	016394;		
DT	01-NOV-1997 (Rel. 35, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	Exostosin-1 (Putative tumor suppressor protein EXT1) (Multiple		
DE	exostosins protein 1).		
GN	EXT1		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxId=9606;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
CC	TISSUE=Placenta;		
RX	MEDLINE=96024648; PubMed=7550340;		
RA	Ahn J., Ludecke H.J., Lindow S., Horton W.A., Lee B., Wagner M.J.,		
RA	Horsthenke B., Wells D.E.;		
RT	"Cloning of the putative tumour suppressor gene for hereditary		
RL	multiple exostosins (EXT1)."		
RL	Nat. Genet. 11:137-143(1995).		
RN	[2]		
RN	VARIANT EXT1 LEU-340.		
RX	MEDLINE=97136573; PubMed=8981950;		
RA	Hecht J.T., Hogue D., Wang Y., Blanton S.H., Wagner M., Strong L.C.,		
RA	Raskind W., Hansen M.F., Wells D.E.;		
RT	Hereditary multiple exostosins (EXT1): mutational studies of familial		
RT	EXT1 cases and EXT-associated malignancies."		
RL	Am. J. Hum. Genet. 60:80-86(1997).		
RN	[3]		
RP	VARIANTS EXT1 ASP-339 AND CYS-340.		
RX	MEDLINE=97465597; PubMed=9326317;		
RA	Philippe C., Porter D.E., Emerton M.E., Wells D.E., Simpson A.H.R.W.,		
RA	Monaco A.P.;		
RT	"Mutation screening of the EXT1 and EXT2 genes in patients with		
RT	hereditary multiple exostosins."		
RL	Am. J. Hum. Genet. 61:520-528(1997).		
RN	[4]		

[illegible]

Query Match	17.9%;	Score 871;	DB 1;	Length 746;
Best Local Similarity	28.7%;	Pred. No. 7.6e-55;		
Matches 236;	Conservative 143;	Mismatches 236;	Indels 206;	Gaps 33

QY	151	PLSLPIRILP----	EKKDGLP--PRPATG-----	-----GLHNCFDYSRCPLNSG	194
QY	151	PLSLPIRILP---- <td>EKKDGLP--PRPATG-----<td>-----GLHNCFDYSRCPLNSG<td>194</td></td></td>	EKKDGLP--PRPATG----- <td>-----GLHNCFDYSRCPLNSG<td>194</td></td>	-----GLHNCFDYSRCPLNSG <td>194</td>	194
DB	56	PRFPRPLRPVPMQOLENEDSVHS	SPQKDAANSIYKKGKCMESC	CFDTLCK-KNGF	1144
QY	195	PVYV--DSDFVEGSYLDLPVKA	FOATARANYVTEMADI	CLYVILVGE	MOEPVLR
DB	115	KYVVYPOQGEKIEASY-----	-QNIILAI	EGSFRTYSDP	SOACTFLVILSDTLDRD-QLS
QY	253	PA---ELRKOLYSLPRHMTGD	HNVIYNL-----SKSDTON	LLYVNSGR	AMVASTFTYV
DB	168	FOYVHNLSKQVOSHLW--N	GRNHLITFLYS	GTMPDYTEDG	FDFI--GQALAKASISTE
QY	307	QYRPGFDLVSPV-----	NAMSEPNEM	EIRPQVVRKYL	FTFTQGEKI-----ESLRS
DB	225	NFRNFDVSI--PLRSKDH	PRTGEGENGFLK--NTI	PLRKALYV	EKKGRYLITLGISDTRNA
QY	357	LOEARSFEENE	GGPRADYDRIITATLK-----	AVODSKI	DYOVLEFTCKNQP
DB	283	LYHNHNGDYV-----	-LTTCKHG	KDMOKHNSD	CDRDNTEY-----
QY	411	EWALGGEREDRELL	ELTKLSTFALLITTPG	DPRLVISSG	ATRLFEAL
DB	320	-----EKYDREH	NHNTFLCY--PKRRRL-----	GSFRLE	ALDAACVPMLSN
QY	471	PYQMLQNE	EAALVYPKPRVTEV	HFLKLS	DSLSDDLARROGR
DB	367	PFSEVINNO	AAVIGDELLQI	ISTISIHODK	LALROOTOFL
QY	531	LAMIFTRI	QIQAAPRI	REDAAEIHR	SGKAGTDP
DB	427	LEITDRI-----	-FKHIS-----	-----	-----
QY	591	RNFUTLV	TVDFRYSWNC	ARGPFLRP	PHTPDPVLR
DB	440	RNSLI-----	-WNKHPG--GLF-----	-VLPOYS	YLADFPYUANGLK
QY	645	GSGRFOA	ALGNGNP	PREFTYV	MLTIERE
DB	477	--PSKEFV	YVIAHVP-----	-LVSSQ	QVQLKLVA
QY	704	EDLLP	DAIGVIMV	RTEKNS	LNRNRELP
DB	526	KH--RMP	AAVAVPVVYIEG	SKYMSR	PLPYDNIIT
QY	764	ARDRLV	GPGRYH	ANDIPHO	QSMLYNSNT
DB	565	PPERIV	GVPARSH	FENDSK	EMKMGYTSKVT
QY	824	WVDEYI	NEDEAM	LVLSHITR	KPPIKYT-----
DB	645	WVDQ	LANKED	ILMFLV	SAVTKLP
QY	868	HDDSH	FERH	KKICINE	FVYKVG
DB	694	-DPDHF	AOROS	CMNTF	ASMF
RESULT	7	EXLI_HUMAN	STANDARD:	PRT:	676 AA.
AC	092935:	15-JUL-1998	(Rel. 36, Created)		
DT	15-JUL-1998	(Rel. 36, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Exostosin-like 1 (Exostosin-L) (Multiple exostosin-like protein).				
GN	EXTL1 OR EXTL.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
NCBI	TextID=9606;				
RN	[1]				
SEQUENCE	FROM N.A.				
rx	MEDLINE=97189339;				
rx	PubMed=9037597;				

```

QY 393 QVLVEFTCKNPKPSLPTFMALCGEREDRLLEKLSTFALITTPGDPRLVSSG---CA 448
Db 252 -----CEQDPGPG-----GTQROETLPNATFCLT-----SGRPEPA 283
QY 449 TRLEFALVGAAPVVLGQVOVLPRYODMLQWNEALVPRVTEVHFLRLSDSLAM 508
Db 284 SRLFOALQAGCIPVLISPRMELPFSEVIDWTKAIVADERLPLQVLAALQEMSPARVAL 343
QY 509 RROGRELMEYESTADSLTFNVLAIRIRIQIPAPRIDEAAAEIPHSKRAKATDPNMA 568
Db 344 RQOQFELMDATFSSVEKYIHTTLEVIQDRI-----FGTS----- 377
QY 569 DNGDLGLSPVETPEPPYASPRYLNRNFTLVTDYFNSMNCAPGPHLPPTPDPVLPSEAK 628
Db 378 -----ANPSLL-----WNSPPGA--LLALSTS--TSPQDP 405
QY 629 FLGSGTGFRPIGGAGSGKEFOALGQNVPEQFTVMLTYEREVLNMSLERLNGLPY 688
Db 406 FYLLQOGSRPEG-----RSALIWGPPQGPPLKILQAVAGSQH 444
QY 689 LNKVYVW-NSPKLPSEDLMPDIGVPIMVVTRTEKNSLNKRFLEPNEIETPAILSIDDA 747
Db 445 CAQTLVMSNRPPLPSR---WPEYAVPLTVIDGHR-KVSDREFFPYSTIRTDAILSIDARS 500
QY 748 HLHDEIMFGFRVWREARDRIYGFPGRYHANDIPHOSWLYNSNYSCELSMVLTGAEPFK 807
Db 501 SLTSEVDFALVWOSPERNAVGLTSHFDEAHGKGITYAKERTNESMVLTTAAEFYHR 560
QY 808 YYAVLYTYVMPQAIRDYNDEVINCEDIANMFLVSHITRKPRIKY-----TSRWTFRCG- 861
Db 561 YYHTLFTHSLPKALRTLADEAPTCVDVLMNFIYAAVTKLPPIKYVKGQGEAPLAPFAG 620
QY 862 -----CPQALSHDSHFERHKCINFEVKYGYMPLLYTORVSVLFK 905
Db 621 PGPFRPPAPAPD-----CINQIAAAGMHPMLLSRLRDPVLEK 660

RESULT 8
EXTL_HUMAN
ID EXT1_HUMAN STANDARD; PRT; 330 AA.
AC Q9UBQ6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Exostosin-like 2 (EXT-related protein 2).
GN EXTL2 OR EXT2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
NC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=98111853;
RA Wuyls W., Van Hul W., Hendrickx J., Speleman F., Wauters J.,
De Boule K., Van Roy N., Van Agtmael T., Bossuyt P., Willems P.J.;
RT "Identification and characterization of a novel member of the EXT gene
family, EXTL2."
RL Eur. J. Hum. Genet. 5:382-389(1997).
RN [2]
RP SEQUENCE FROM N.A.
MEDLINE=98139867; PubMed=9473480;
RA Saito T., Seki N., Yamauchi M., Tsuji S., Hayashi A., Kozuma S.,
Hori T.-A.;
RT "Structure, chromosome location, and expression profile of EXT1 and
EXT2, new members of the multiple exostos gene family."
RL Biochem. Biophys. Res. Commun. 243:61-66(1998).
CC -|- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
reticulum (By similarity).
CC -|- TISSUE SPECIFICITY: UBIQUITOUS.
CC -|- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF000416; AAC02898.1; -
DR EMBL: AB009284; BAA24081.1; -
DR MIM: 602411; -
DR InterPro: IPR004263; Exostosin.
DR Pfam: PF03016; Exostosin; 1.
KW Multigene family; Transmembrane; Signal-anchor.
FT TRANSMEM 23 43 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE .330 AA; 37465 MW; 6976BE7EC6F588C8 CRC64;
SQ
Query Match 7.5%; Score 367; DB 1; Length 330;
Best Local Similarity 30.8%; Pred. No. 4.2e-19;
Matches 86; Conservative 49; Mismatches 98; Indels 46; Gaps 7;
QY 656 GNVPEQFTVMLTYEREVLNMSLERLNGLPYLNKVVVWNSPKLPSEDLMPDIG--- 712
Db 59 GKSTMDSTLLIMQTYNRKIDLLKLNHYQAVPNLAKYLVANNNGEKAPDELMNSLGPH 118
QY 713 VPINVYTRTEKNSLNKRFLEPNEIETAILSIDDAHLHDEIMFGFRVWREARDRIYGF 772
Db 119 IPIVFCQGTANRRMRRLQVPELEFNALVAVDDPLISTPPLVFAFSVWQGFPGIYGFV 178
QY 773 GRYHAMDIPHOSWLYN-----SNYSCELSMVLTGAEPFKRYAYLYKSYVMPQAIR 822
Db 179 PRKH---VSTSSGITYSGSFEMQAPGSGNGDQYSNVLIGASFNSKYLELPQR-QPAVH 234
QY 823 DMVDEVINCEDIANMFLVSHITRKPRIKY-----TSRWTFRCGCPQA 865
Db 235 ALIDDTQNCDDIANMFIATKIHGTSGLFYKPVVMMNDLEKFTNGSYGSMW----- 284
QY 866 LSHDSHFERHKCINFEVKYGYMPLLYTORVSVLFK 904
Db 285 --HRAEHALGRSCYNKLVNIDSMPLRYSNIMISQGF 321

RESULT 9
EXTL_CAEEL
ID EXT1_CAEEL STANDARD; PRT; 378 AA.
AC O01704;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Multiple exostos homolog 1.
GN RIB-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
NC Rhabditidae; Pelodirinae; Caenorhabditis.
NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=97264342; PubMed=9110175;
RA Cline G.A., Ashley J.A., Shah S., Lovett M.;
RT "The structure of the human multiple exostos 2 gene and
characterization of homologs in mouse and Caenorhabditis elegans."
RL Genome Res. 7:359-367(1997).
CC -|- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```


DE V-type ATP synthase subunit I (EC 3.6.3.14) (V-type ATPase subunit I).
 GN API OR AF1159.
 OS Archaeoglobus fulgidus.
 CC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
 CC Archaeoglobus.
 OX NCBI_TaxID=2234;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
 RX MEDLINE=98049343; PubMed=9389475;
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
 RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
 RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
 RA Fleischman R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
 RA Kirschner E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
 RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
 RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
 RA Cotton M.D., Spriggs T., Arlatch P., Kaine B.P., Sykes S.M.,
 RA Sadow P.M., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
 RA Venter J.C.;
 RT "The complete genome sequence of the hyperthermophilic, sulphate-
 RT reducing archaeon Archaeoglobus fulgidus.";
 RL Nature 390:364-370(1997)
 CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
 CC GRADIENT ACROSS THE MEMBRANE.
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
 CC H(+) (out).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE001023; AAB90076.1; -;
 DR TIGR: AF1159; -;
 DR InterPro: IPR002490; V_ATPase_sub_a.
 DR Pfam: PF01496; V_ATPase_sub.a: 1.
 DR KW Hydrolyase; Hydrogen ion transport; Transmembrane; Complete proteome.
 FT TRANSMEM 341 361 POTENTIAL.
 FT TRANSMEM 390 410 POTENTIAL.
 FT TRANSMEM 449 469 POTENTIAL.
 FT TRANSMEM 490 510 POTENTIAL.
 FT TRANSMEM 538 558 POTENTIAL.
 FT TRANSMEM 564 584 POTENTIAL.
 FT TRANSMEM 590 610 POTENTIAL.
 FT TRANSMEM 617 637 POTENTIAL.
 SQ SEQUENCE 676 AA; 76649 MW; E2E0AC8890C9CA45 CRC64;

Query Match 2.3%; Score 113; DB 1; Length 676;
 Best Local Similarity 21.9%; Pred. No. 2.3; Indels 80; Gaps 15;
 Matches 68; Conservative 39; Mismatches 124;

OY 259 -QLY-----SLPHWRTDGHNVIIIMLSRKSPTONILY- NVSTGRAMVA---QSTFYVQYR 309
 Db 417 YQHIFISQLEHLXEFHGHGHPFVRVEMGKILLFATFYIGAKILFGALGRTNYNVE 476
 OY 310 PGF-DLVVSL 319
 Db 477 HGLKDAILEKL 487
 RESULT 12
 CENE_HUMAN STANDARD; PRT; 2663 AA.
 AC Q02224;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Centromeric protein E (CENP-E protein).
 GN CENPE.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93024922; PubMed=1406971;
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 RT mitosis.";
 RL Nature 359:536-539(1992).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=95196755; PubMed=7889940;
 RA Throver D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
 RT microtubule motor.";
 RL EMBO J. 14:918-926(1995).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 RT interactions with the kinetochore proteins CENP-F and hBURL1.";
 RL J. Cell Biol. 143:49-63(1998).
 CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
 CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
 CC AND/OR SPINDLE ELONGATION.
 CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
 CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z15005; CAA78727.1; -;
 DR PIR: S28261; S28261.
 DR HSSP: P17119; 3KAR.
 DR MIM: 117143; -;
 DR InterPro: IPR001752; kinesin.
 DR Pfam: PF00225; kinesin; 1.
 DR PRINTS: PR00380; KINESINHEAVY.
 DR SMART: SM00129; KISC; 1.
 DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.
 KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;

Cell cycle: Centromere.
 KW DOMAIN 1 335 KINESIN-MOTOR.
 FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
 FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
 FT NP_BIND 86 93 ATP (BY SIMILARITY).
 SQ SEQUENCE 2663 AA: 312087 MW: 69FC13880C8C8C8B CRC64:

Query Match 2.3%; Score 113; DB 1; Length 2663;
 Best Local Similarity 18.4%; Pred. No. 18;
 Matches 143; Conservative 114; Mismatches 308; Indels 214; Gaps 33;

54 LBDADAGKIFGPRVGNELCEVKNVLDLCRIEVSSELLQLEAKROELNSTAKLNK 113
 1877 LHNLEEMKSVKMER--DNLRLVEETLKL--EKDOLKESIQETKARLEIQOELKTARML 1932
 114 IEACKSIENAKODL--LQKNVISQTEHSYKELMAONQPKLSPIRLPEKPDAGLP 169
 1933 SKHKETVQDLRKIKISEKTIQISDIOKDLKSKDELQKIQOELQKKELOLRKVEDYNMS 1992
 170 PPKATGCRRLNCFDYSRCPETSGFPVYV--YDSQFVFGSYLDPLVKAQATARANY 227
 1993 HKKINEMEQKKQFE-----PVLCKCEMDN-----QLTKLHESLEIRIV 2035
 228 VTENAD-----TACLYVILVGENQEPVYLRPALEKQVSLPHMRDGHN 272
 2036 AKERDELRRIKESLKMEROQFIATLRERIMARDQNHQV--KP--EKRLLS-----DGOQ 2085
 273 HVILNLSRK--SDQNLLYNVSTRANVAOSTFYTVQVRGFDLVSLVHAMSEPNMEI 331
 2086 HLMESVREKSRKIKELKRY-----EMDDHCECLNRLSLD-----EKEIEFHRI 2131
 332 PPOVPKRYKLTFFQGEKTESLRSQLEARSFEEMEG-----DPADYDDRRI 380
 2132 MKRLKTVLVSVTKIKEQHCINKFEMD--FIDVEKQKQKELLIKIQHQQDDVPSREL 2188
 381 ATKAVO--DSKIDQVLEVETCKNOKPKSLPTW--ALGGER-----DRIELLK 426
 2189 RDKLQNMMDLHEELIKDPS--ESEFPKITEFOQVLSNRKEMTOFLEWMLNTRFIDEK 2246
 427 LSTFALITFGDPRLVYSSGCATRLP--EALVGAVPVLEQOVLPRQDMQWNE-- 480
 2247 LKNG--IQKENDRICOVNFFNNRITAINNESTEFERBSATISKEMQDLKSLKNEKL 2304
 481 -----AALVVP-----KPRVT-----EVHFLSLSDSLAMRQ 512
 2305 FKNVQTLKTSLSAGQOVNPTQDNKNPHVSRATQLTETIKIRELNSLHAKESAMHKS 2364
 513 RPLMETYESTADSIPTNTVLAMIRTIQIPAPRIEAAAEIPHRSGK----- 559
 2365 KII--KMQELEVTNDIILKQAKV-----HESNCKLEKTKETIQVLQ 2405
 560 ---AAGDPMANNGDLDGPPV-----EPPEYASPRYLNFILTVDFY 601
 2406 DKVALGAKPKKEIEDKMKLVKIDLEKKNNAKEPEKEISATATAYEYQVIRLLENL 2465
 602 RSNMCAKGPFRHLPFRPFDPVLPSEAKFLGSGTGFRPI--GGGAGSGKFGQALGQNP 659
 2466 RRSQAOQDTSVISEHT--DP-QPSN-----KPLTGGGSGIQQNKKALIKSEHI 2512
 660 REOPTVVMLTREVEVLMSLERLNGLPYANKVYVW-----NSRKL 702
 2513 RLEKEISKLKQONQEOILKQKNELLNNQHLNSNEVKTWKERTLRKAHKQVTCENSPKSP 2571

RESULT 13
 ID DYHC_CAEEL STANDARD; PRT; 4568 AA.
 AC 019020;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE dynein heavy chain, cytosolic (DYHC).

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditioidea;
 OC Rhabditidae; Peloderae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=96114101; PubMed=8674131;
 RA Lye R.J., Wilson R.K., Waterston R.H.;
 RT "Genomic structure of a cytoplasmic dynein heavy chain gene from the
 nematode Caenorhabditis elegans."
 RL Cell Motil. Cytoskeleton 32:26-36(1995).
 CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
 MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
 ORGANELLES ALONG MICROTUBULES.
 CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
 INTERMEDIATE AND LIGHT CHAINS.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L33260; AAC37251.1; -
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR004273; Dynein_heavy.
 DR Pfam: PF03028; Dynein_heavy; 1.
 DR SMART: SM00382; AAA; 1.
 DR Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
 FT DOMAIN 587 652
 FT DOMAIN 814 844
 FT DOMAIN 1241 1274
 FT DOMAIN 1324 1340
 FT DOMAIN 1559 1591
 FT DOMAIN 1964 1992
 FT DOMAIN 3132 3229
 FT DOMAIN 3339 3432
 FT DOMAIN 3707 3739
 FT DOMAIN 4359 4386
 FT NP_BIND 1865 1872
 FT NP_BIND 2163 2170
 FT NP_BIND 2537 2544
 FT NP_BIND 2880 2887
 SQ SEQUENCE 4568 AA: 521568 MW: 028E52684F361676 CRC64;

Query Match 2.3%; Score 112.5; DB 1; Length 4568;
 Best Local Similarity 18.7%; Pred. No. 42;
 Matches 158; Conservative 96; Mismatches 256; Indels 335; Gaps 36;

87 ESYSEELLQLEAKROELNSTAKLNKIEACKSIENAKODLLQKNVISQTEHSYKELM 146
 3353 EPLRELKRLQEPAAKKTQGVKVDVIRTELSEIGKYKEVYQAL--IQAEINIKODLL 3409
 147 AQONPKLSLPIRLPEKPDAGLPKPKATRCGRRLNCFDYSRCPVTSFPYVYDSQFV- 205
 3410 SYOE-----RVNRSTELLSLRSEDRKWSGSGAFSGOQMSVLG 3448
 206 -----FGSYLDPLVKAQ-----FQATARANYVTENADIACL--YVILVGENQEPV 249
 3449 DALLSNFLAYAGYVQDMRLDELFFHKFNHVVNAGLHFRH--DLAREVYSTVDD----- 3501
 250 VLKPALEKQVSLPHMRDGHNVIIINLSRKSDTONLLVNVSTGRAMVAOSTFYTVQVR 309
 3502 -----RLQWQNLNSLPVDDLCTENAIMLHREN----- 3528
 310 PGFDLVVSPVYHAMSEPNMEIIPQVPVKKRYL-FFGGKI-----ESLRSIQEAR 361

FT	DOMAIN	742	790	LAMININ BGF-Like 6.
FT	DOMAIN	791	848	LAMININ BGF-Like 7.
FT	DOMAIN	849	901	LAMININ BGF-Like 8.
FT	DOMAIN	902	950	LAMININ BGF-Like 9.
FT	DOMAIN	951	997	LAMININ BGF-Like 10.
FT	DOMAIN	998	1043	LAMININ BGF-Like 11.
FT	DOMAIN	1044	1089	LAMININ BGF-Like 12.
FT	DOMAIN	1090	1149	LAMININ BGF-Like 13.
FT	DOMAIN	1150	1159	LAMININ BGF-Like 14.
FT	DOMAIN	1160	1361	LAMININ BGF-Like 14 (N-TERMINAL).
FT	DOMAIN	1362	1402	LAMININ BGF-Like 14 (DOMAIN IV A).
FT	DOMAIN	1403	1451	LAMININ BGF-Like 14 (C-TERMINAL).
FT	DOMAIN	1452	1508	LAMININ BGF-Like 15.
FT	DOMAIN	1509	1555	LAMININ BGF-Like 16.
FT	DOMAIN	1556	2116	LAMININ BGF-Like 17.
FT	DOMAIN	2117	2297	LAMININ G-Like 1.
FT	DOMAIN	2305	2481	LAMININ G-Like 2.
FT	DOMAIN	2486	2673	LAMININ G-Like 3.
FT	DOMAIN	2713	2885	LAMININ G-Like 4.
FT	DOMAIN	2890	3070	LAMININ G-Like 5.
FT	DOMAIN	1706	1796	COILED COIL (POTENTIAL).
FT	DOMAIN	1968	1989	COILED COIL (POTENTIAL).
FT	DOMAIN	2088	2120	CELL ATTACHMENT SITE.
FT	DOMAIN	2534	2536	CELL ATTACHMENT SITE.
FT	DOMAIN	270	279	BY SIMILARITY.
FT	DOMAIN	272	290	BY SIMILARITY.
FT	DOMAIN	292	301	BY SIMILARITY.
FT	DOMAIN	297	305	POTENTIAL.
FT	DOMAIN	304	324	BY SIMILARITY.
FT	DOMAIN	327	336	BY SIMILARITY.
FT	DOMAIN	329	361	BY SIMILARITY.
FT	DOMAIN	364	373	BY SIMILARITY.
FT	DOMAIN	376	394	BY SIMILARITY.
FT	DOMAIN	397	409	BY SIMILARITY.
FT	DOMAIN	427	438	BY SIMILARITY.
FT	DOMAIN	439	441	BY SIMILARITY.
FT	DOMAIN	441	451	BY SIMILARITY.
FT	DOMAIN	454	467	BY SIMILARITY.
FT	DOMAIN	471	471	BY SIMILARITY.
FT	DOMAIN	473	482	BY SIMILARITY.
FT	DOMAIN	485	500	BY SIMILARITY.
FT	DOMAIN	742	751	BY SIMILARITY.
FT	DOMAIN	744	757	BY SIMILARITY.
FT	DOMAIN	760	769	BY SIMILARITY.
FT	DOMAIN	772	788	BY SIMILARITY.
FT	DOMAIN	791	806	BY SIMILARITY.
FT	DOMAIN	793	816	BY SIMILARITY.
FT	DOMAIN	819	828	BY SIMILARITY.
FT	DOMAIN	831	846	BY SIMILARITY.
FT	DOMAIN	849	863	BY SIMILARITY.
FT	DOMAIN	851	870	BY SIMILARITY.
FT	DOMAIN	873	882	BY SIMILARITY.
FT	DOMAIN	885	899	BY SIMILARITY.
FT	DOMAIN	902	914	BY SIMILARITY.
FT	DOMAIN	904	921	BY SIMILARITY.
FT	DOMAIN	923	932	BY SIMILARITY.
FT	DOMAIN	935	943	BY SIMILARITY.
FT	DOMAIN	951	963	BY SIMILARITY.
FT	DOMAIN	953	969	BY SIMILARITY.
FT	DOMAIN	971	980	BY SIMILARITY.
FT	DOMAIN	983	995	BY SIMILARITY.
FT	DOMAIN	998	1007	BY SIMILARITY.
FT	DOMAIN	1000	1014	BY SIMILARITY.
FT	DOMAIN	1016	1025	BY SIMILARITY.
FT	DOMAIN	1028	1041	BY SIMILARITY.
FT	DOMAIN	1044	1056	BY SIMILARITY.
FT	DOMAIN	1046	1063	BY SIMILARITY.
FT	DOMAIN	1065	1074	BY SIMILARITY.
FT	DOMAIN	1077	1087	BY SIMILARITY.
FT	DOMAIN	1403	1412	BY SIMILARITY.
FT	DOMAIN	1405	1419	BY SIMILARITY.
FT	DOMAIN	1422	1431	BY SIMILARITY.
FT	DOMAIN	1434	1449	BY SIMILARITY.

FT	DISULFID	1452	1466	BY SIMILARITY.
FT	DISULFID	1454	1476	BY SIMILARITY.
FT	DISULFID	1479	1488	BY SIMILARITY.
FT	DISULFID	1491	1506	BY SIMILARITY.
FT	DISULFID	1509	1521	BY SIMILARITY.
FT	DISULFID	1511	1528	BY SIMILARITY.
FT	DISULFID	1530	1539	BY SIMILARITY.
FT	DISULFID	1542	1553	BY SIMILARITY.
FT	DISULFID	1556	1556	INTERCHAIN (PROBABLE).
FT	DISULFID	1560	1560	INTERCHAIN (PROBABLE).
FT	CARBOHYD	38	38	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	555	555	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	665	665	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	763	763	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	926	926	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	952	952	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1045	1045	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1407	1407	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1579	1579	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1596	1596	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1678	1678	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1689	1689	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1698	1698	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1717	1717	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1804	1804	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1894	1894	N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 2.3%; Score 111.5; DB 1; Length 3075;
 Best Local Similarity 20.0%; Pred. No. 28;
 Matches 119; Conservative 105; Mismatches 237; Indels 135; Gaps 27;

QY	71	NELCENKVID-LCRIRSEVSEBLLDLKARQELNSELATLNLKIEACKSIENAKODLL	129
DB	1717	NATLELKAAADLLSQIOENYOKPLEELEVLEKAAASHVLKSHNNELKAAEALVBEAEAKMQ	1776
QY	130	QLKNVI-----SOTHEKYLMAONQKLSPIRLPEKXDAGLPPKATRGRLHNCFD	164
DB	1777	ESNLLMLMANANREFSDKLLHVOEONLTVSELV-----OGGLIDAAAOOTDAVADALE	1632
QY	185	YSRCPITSGPPVYVYDDQFVFGS-----YLDPLVKQAFQATARANYVTEN--ADIACL	237
DB	1833	HLE-----DHQDKLLMSAKIRHHIDLVNMSQRNVDLYRAEDHATERQRL	1861
QY	238	YVILVGMQEPVVLRAPELEKOLYSLPHMRTDG---HNHVIINLSKRSPTQNLN--V	231
DB	1882	ADVLVSGLEN---IRNVSLSNATSAAYVHYNIOSLIESEBELARDHARTVETSLSESIV	1938
QY	292	STGRANVAOSTFYTVQVR-----PGFDLVVSPVHAMS--EPFMFIPO-----	334
DB	1939	SNGRAAVORSRFLKEBNNLSRKLPGTALSELURNKTNFQENAVITQTQNESLILR	1998
QY	335	-VP-----VKRYLTFQGEKIESLRSSLOEARSPEEMEGDPADYDRIATLK--	384
DB	1999	ALPEGIRDKAKTKRELAT---SASQSAVSTLRDVAAGSLQSLNMTAS--LSRVNTTLRET	2053
QY	385	-AAVDSKLDQVLEFTCKN-QPKPSLPTEMALCGEDRDLBELKSTFLIITPGDPLR	441
DB	2054	HOLDODSTMAITLLAGRKVKKLEDEIOANL-----LFDRLKPLKLEENLSRNLSEIKL	2104
QY	442	VISGCAITLFEPLLEVGAADVVLGEQVQLPYODML---QNNEAALVYPRKRVTEVHLLR	498
DB	2105	LISO--ARKQASATKA---VSADRCICIRAYQPOISSTNYNTLTLANKTQEPDNLFLYLG	2159
QY	499	SLSDDLLAMR-ROGR--FLMETVYFSTADSIPTVLAMITRIQIAPAPIREEAAEI--	553
DB	2160	SSTASDPLAIVEMRGRVAFPLMD-----LGSSTRLFEFPFPIDDNKNWHSIHV	2206
QY	554	-----PHSGKAAGTDPNMAONGDLDL-----GPVEEP	582
DB	2207	ARRGNIGLSLVKEMSNOKSPRTSKSPGT-ANVLDVNNSTLMPVGLGGOIKKSP	2261
RESULT	15		

CPXR_BRAJA
ID CPXR_BRAJA STANDARD: PRT: 429 AA.
AC 059204;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Cytochrome P450 Bf-3 (EC 1.14.14.-) (Cytochrome P450 114).
GN CYP114.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-USA 110;
RA Tully R.E., Keister D.L.;
RT "Cloning and mutagenesis of a cytochrome P-450 locus from
RT Bradyrhizobium japonicum that is expressed anaerobically and
RT symbiotically."
RT Appl. Environ. Microbiol. 59:4136-4142(1993).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN-USA 110;
RX MEDLINE=98322110; PubMed=9655913;
RA Tully R.E., van Berkum P., Loyins K.W., Keister D.L.;
RT "Identification and sequencing of a cytochrome P450 gene cluster from
RT Bradyrhizobium japonicum."
RL Biochim. Biophys. Acta 1398:243-255(1998).
CC -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U12678; AAC28890.1; ALT_INIT.
DR HSSP; 000441; 10XA.
DR InterPro: IPR001128; Cyt_P450.
DR Pfam; PF00067; P450.1.
DR PRINTS; PR00359; BP450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
DR Oxidoreductase; Monooxygenase; Electron transport; Heme.
DR BINDING 376 376 HEME (BY SIMILARITY).
DR SEQUENCE 429 AA; 46920 MW; B4CC31406007983 CRC64;

Query Match 2.3%; Score 110.5; DB 1; Length 429;
Best Local Similarity 23.7%; Pred. No. 1.8;
Matches 94; Conservative 45; Mismatches 148; Indels 109; Gaps 17;
QY 310 GFEPLVSPVLAHSENFMEIPPOVYKRYLFTFOGKTESLRSSLOEARSEFEEMEG 369
DB 85 PSLMLASTL--AMKDP-----PTHTRLRLISRFTMGQIDNLRPSIA----- 126
QY 370 DPPADYDRIATLKAQVDSKLDQVIVFTCKNPKPSLPTFEMALGCRERDLKLIST 429
DB 127 -----RIVA-----ARLDGI-----TPALERGEAVDLHREFALAPMLVF 161
QY 430 FALITTPGDPRLVYSSGATRLFEALEGAVPVV-----GEVOQLPYQDMLQWNEAL 483
DB 162 AELFGMPQDDMFELAAIGT-ILEGLPHASDPQLAADAASARVQAYFGDLIQKRT-- 218
QY 484 VPKPRVTEVHFL-----RSLSDSLAMRRQGRFLMET-----YESTADSIPTVLA 532
DB 219 ---DPRDIYSMLVGAHDDADTLSDAELISM-----LWCMILGFTVTAASIDHAYLA 269
QY 533 MI-----RTRIQIPAAPIREBAAEIPHRSGKAAGTDPNMADNCDLGLGVETEPYASP 587

DB 270 MLAYPEQRHWLQADAAARAAVEEVLRCDAAPAMFESSIRIQR-DIELGV-VIRKNADV 327
QY 588 RYLRFETLTVTDFYRSMNCAPGPFHLFPHTPPDPVLPSEAKFLGSGTFRPIGGAGSG 647
DB 328 RVL-----IASGNRDPDAF-----ADPRFPAPRYGTSPGMSIDGKIMLSFG 370
QY 648 KEQALAGNVPREQFTVYVMTYEREVLMNSLERL 683
DB 371 HGTHFCLAGAQLARVQ-----LAESLPRI 393

Search completed: September 4, 2002, 15:03:02
Job time: 266 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 4, 2002, 14:57:41 ; Search time 37.88 Seconds

(without alignments)
4196.999 Million cell updates/sec

Title: US-09-809-920-4

Perfect score: 4873

Sequence: 1 MTGYTMLRNGAGNGGCTCM.....DSVLEKTRLPMDKTCFKFI 919

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPREMBL_19:*

- 1: sp_Archaea:*
- 2: sp_Bacteria:*
- 3: sp_Fungi:*
- 4: sp_Human:*
- 5: sp_Invertebrate:*
- 6: sp_Mammal:*
- 7: sp_Mhc:*
- 8: sp_Organelle:*
- 9: sp_Phage:*
- 10: sp_Plant:*
- 11: sp_Rodent:*
- 12: sp_Virus:*
- 13: sp_Vertebrate:*
- 14: sp_Unclassified:*
- 15: sp_Virus:*
- 16: sp_Bacteriophage:*
- 17: sp_Archaeo:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4755	97.6	919	11	Q9JMA8
2	4720.5	96.9	918	11	Q9WVL6
3	2314	47.5	972	5	Q9XZ08
4	1252.5	25.7	814	5	O17920
5	1003	20.6	718	11	Q923D6
6	993	20.4	718	6	Q77783
7	939	19.3	728	4	O15288
8	870.5	17.9	746	4	Q9BVR9
9	870.5	17.9	746	11	Q9JKB2
10	841	17.3	717	5	Q91B9
11	832	17.1	738	13	Q98SV5
12	796.5	16.3	760	5	O76796
13	795.5	16.3	760	5	Q9V730
14	639	13.1	669	11	Q9JUV7
15	586	12.0	131	11	Q9D7G5
16	379.5	7.8	330	11	Q9ES89

17	372.5	7.6	329	11	Q9CX90
18	360	7.4	334	10	Q9LY62
19	312.5	6.4	741	10	Q9AUK5
20	306.5	6.3	764	10	Q9L275
21	288	5.9	329	10	Q9C975
22	251.5	5.2	382	5	Q9U3J6
23	211	4.3	67	6	Q9SKP6
24	153	3.1	129	4	Q00245
25	135.5	2.8	498	10	Q9FEB4
26	129	2.6	1310	10	Q949K0
27	127.5	2.6	341	10	Q94AA9
28	123.5	2.5	425	10	Q9SZK4
29	123	2.5	1568	11	Q9ESK9
30	121.5	2.5	415	10	Q94Q08
31	121.5	2.5	511	10	Q9LFL2
32	121.5	2.5	1185	4	Q13135
33	121.5	2.5	1202	4	Q13136
34	120.5	2.5	1257	4	Q75334
35	120	2.5	460	10	Q82181
36	120	2.5	893	16	Q9X1G2
37	120	2.5	1267	4	Q75145
38	119.5	2.5	824	13	Q90778
39	119	2.4	846	4	Q75130
40	118.5	2.4	412	10	Q9F2J1
41	117.5	2.4	811	13	Q9Y1B9
42	117	2.4	440	10	Q64805
43	117	2.4	1331	17	Q9YEG9
44	116	2.4	499	10	Q9S565
45	114.5	2.3	1024	10	Q23517

ALIGNMENTS

RESULT	ID	Q9JMA8	PRELIMINARY:	PRT:	919 AA.
Q9JMA8	Q9JMA8	Q9JMA8			
DT	01-OCR-2000 (TREMBLrel. 15, Created)				
DT	01-OCR-2000 (TREMBLrel. 15, Last sequence update)				
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)				
DE	REG RECEPTOR.				
GN	REG.				
OS	Rattus norvegicus (Rat).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
OX	NCBI_Taxid=10116;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-WISTAR; TISSUE-PANCREATIC ISLETS;				
RX	MEDLINE=20219101; PubMed=10753661;				
RA	Kobayashi S., Akiyama T., Naka K., Abe M., Tajima M., Shervani N.J.,				
RA	Uno M., Matsuno S., Sasaki H., Takasawa S., Okamoto H.;				
RT	"Identification of a receptor for reg (regenerating gene) protein, a				
RT	pancreatic beta-cell regeneration factor."				
RL	J. Biol. Chem. 275:10723-10726(2000).				
DR	EMBL: AB033367; BAA92895.1; -				
KM	Receptor.				
SQ	SEQUENCE 919 AA; 104443 MW; 6F42EA80FD5C2807 CRC64;				

Query Match 97.6%; Score 4755; DB 11; Length 919;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 894; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

Qy	1	MTGYTMLRNGAGNGGCTCMRMSNRIRLTWSTLPIVLPFLINHYITTTIDEADEA 60
Db	1	MTGYTMLRNGAGNGGCTCMRMSNRIRLTWSTLPIVLPFLINHYITTTIDEADEA 60
Oy	61	GKRFPGPVGNELCEVHVLDLCRIRESVSEELQLEAKROELNSETAKLNLKTEACKS 120
Db	61	GKRFPGPVGNELCEVHVLDLCRIRESVSEELQLEAKROELNSETAKLNLKTEACKS 120

```

QY 121 IENAKODLLOLKNVTSOTESHYKELMAONOKLSLPIRLPEKDDAGLPPKATRGCRH 180
Db 121 IENAKODLLOLKNVTSOTESHYKELMAONOKLSLPIRLPEKDDAGLPPKATRGCRH 180
QY 181 NCFDYSRCPLTSGFPVYVYDSDFVGSYLDPLVKAQFQATARANVYTEMADIACLYVI 240
Db 181 NCFDYSRCPLTSGFPVYVYDSDFVGSYLDPLVKAQFQATARANVYTEMADIACLYVI 240
QY 241 LVGENOEPPVLRPALEKOLYSLPHWRTDGHNVYIINLSRKSDDONLNYSTGRANVAQ 300
Db 241 LVGENOEPPVLRPALEKOLYSLPHWRTDGHNVYIINLSRKSDDONLNYSTGRANVAQ 300
QY 301 STEFTVOYRPGFDLVVSPVLAHAMSEPNMEIPPOVPKRYKLTFTQGEKIESLSLOEA 360
Db 301 STEFTVOYRPGFDLVVSPVLAHAMSEPNMEIPPOVPKRYKLTFTQGEKIESLSLOEA 360
QY 361 RSFEEMGDPADYDDRIIATLKAVODSKLDQVLEFTCKNOKPSPSLPTMALCGERED 420
Db 361 RSFEEMGDPADYDDRIIATLKAVODSKLDQVLEFTCKNOKPSPSLPTMALCGERED 420
QY 421 RLELLKSTFALLITTPGDPRLVSSGATRLFEALEVAVPVVIGEYQVLPYODMLQWNE 480
Db 421 RLELLKSTFALLITTPGDPRLVSSGATRLFEALEVAVPVVIGEYQVLPYODMLQWNE 480
QY 481 AALVVPKRYTEVHFHFLRSLSDSLMLAMRROGRFLMETESTADSIINTVLAIRTRIOI 540
Db 481 AALVVPKRYTEVHFHFLRSLSDSLMLAMRROGRFLMETESTADSIINTVLAIRTRIOI 540
QY 541 PAADIREBAAAEIPHRSGKAAGTDPNMAANGDLDLGEVETEPYASPRYLNFLLTYTDF 600
Db 541 PAADIREBAAAEIPHRSGKAAGTDPNMAANGDLDLGEVETEPYASPRYLNFLLTYTDF 600
QY 601 YRSWNCAPGPHLEPHRPPDVLPSSEAKFLSGTGFRPIGGAGSGSKEFOALGCVNR 660
Db 601 YRSWNCAPGPHLEPHRPPDVLPSSEAKFLSGTGFRPIGGAGSGSKEFOALGCVNR 660
QY 661 EOEFTVVALTYEREVEVLNLSLERLNGLPYLNKVVVWVNSPKLPSEDLLMPDIGVIMVYRT 720
Db 661 EOEFTVVALTYEREVEVLNLSLERLNGLPYLNKVVVWVNSPKLPSEDLLMPDIGVIMVYRT 720
QY 721 EKNSLNRRFLPWNIEETEAISIDDDAHLRDEIMFGFRVWREARDRIYVGPGRYHAMD 780
Db 721 EKNSLNRRFLPWNIEETEAISIDDDAHLRDEIMFGFRVWREARDRIYVGPGRYHAMD 780
QY 781 PHOSWLVNSYSCELSVNLGAAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFV 840
Db 781 PHOSWLVNSYSCELSVNLGAAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFV 840
QY 841 SHITRKPIKVTSMWTRCPCGCPQALSHDSDSHHERHKCINFEVYKYGWMLLYTOFRVD 900
Db 841 SHITRKPIKVTSMWTRCPCGCPQALSHDSDSHHERHKCINFEVYKYGWMLLYTOFRVD 900
QY 901 SVLEKTRLPDHDKTCFKFI 919
Db 901 SVLEKTRLPDHDKTCFKFI 919

```

```

RT "Molecular cloning of the mouse homolog of ExTL."
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF083550; AAD42040.1; -.
DR MGI: MGI:1860765; ExTL3.
DR InterPro: IPR004263; Exostosin.
DR Pfam: PF03016; Exostosin; 1.
SQ SEQUENCE 918 AA; 104547 MW; 4EEFEC86E7E6027E CRC64;

Query Match          96.9%; Score 4720.5; DB 11; Length 918;
Best Local Similarity 97.0%; Pred. No. 0;
Matches 891; Conservative 7; Mismatches 20; Indels 1; Gaps 1;

QY 1 MTGYTMLRNGAGNGGQTCMLRWSNRIRLTMLSFTLVLYVEFFLIANYLTTTDEADEA 60
Db 1 MTGYTMLRNGAGNGGQTCMLRWSNRIRLTMLSFTLVLYVEFFLIANYLTTTDEADEA 60
QY 61 GKRIFGPVGWELCEVKVAVLDLCIRSVSEBLLQLEAKROELNSETAKMLKTEACKS 120
Db 61 GKRIFGPVGWELCEVKVAVLDLCIRSVSEBLLQLEAKROELNSETAKMLKTEACKS 120
QY 121 IENAKODLLOLKNVTSOTESHYKELMAONOKLSLPIRLPEKDDAGLPPKATRGCRH 180
Db 121 IENAKODLLOLKNVTSOTESHYKELMAONOKLSLPIRLPEKDDAGLPPKATRGCRH 180
QY 181 NCFDYSRCPLTSGFPVYVYDSDFVGSYLDPLVKAQFQATARANVYTEMADIACLYVI 240
Db 181 NCFDYSRCPLTSGFPVYVYDSDFVGSYLDPLVKAQFQATARANVYTEMADIACLYVI 240
QY 241 LVGENOEPPVLRPALEKOLYSLPHWRTDGHNVYIINLSRKSDDONLNYSTGRANVAQ 300
Db 241 LVGENOEPPVLRPALEKOLYSLPHWRTDGHNVYIINLSRKSDDONLNYSTGRANVAQ 300
QY 301 STEFTVOYRPGFDLVVSPVLAHAMSEPNMEIPPOVPKRYKLTFTQGEKIESLSLOEA 360
Db 301 STEFTVOYRPGFDLVVSPVLAHAMSEPNMEIPPOVPKRYKLTFTQGEKIESLSLOEA 360
QY 361 RSFEEMGDPADYDDRIIATLKAVODSKLDQVLEFTCKNOKPSPSLPTMALCGERED 420
Db 361 RSFEEMGDPADYDDRIIATLKAVODSKLDQVLEFTCKNOKPSPSLPTMALCGERED 420
QY 421 RLELLKSTFALLITTPGDPRLVSSGATRLFEALEVAVPVVIGEYQVLPYODMLQWNE 480
Db 421 RLELLKSTFALLITTPGDPRLVSSGATRLFEALEVAVPVVIGEYQVLPYODMLQWNE 480
QY 481 AALVVPKRYTEVHFHFLRSLSDSLMLAMRROGRFLMETESTADSIINTVLAIRTRIOI 540
Db 481 AALVVPKRYTEVHFHFLRSLSDSLMLAMRROGRFLMETESTADSIINTVLAIRTRIOI 540
QY 541 PAADIREBAAAEIPHRSGKAAGTDPNMAANGDLDLGEVETEPYASPRYLNFLLTYTDF 600
Db 541 PAADIREBAAAEIPHRSGKAAGTDPNMAANGDLDLGEVETEPYASPRYLNFLLTYTDF 600
QY 601 YRSWNCAPGPHLEPHRPPDVLPSSEAKFLSGTGFRPIGGAGSGSKEFOALGCVNR 660
Db 601 YRSWNCAPGPHLEPHRPPDVLPSSEAKFLSGTGFRPIGGAGSGSKEFOALGCVNR 660
QY 661 EOEFTVVALTYEREVEVLNLSLERLNGLPYLNKVVVWVNSPKLPSEDLLMPDIGVIMVYRT 720
Db 661 EOEFTVVALTYEREVEVLNLSLERLNGLPYLNKVVVWVNSPKLPSEDLLMPDIGVIMVYRT 720
QY 721 EKNSLNRRFLPWNIEETEAISIDDDAHLRDEIMFGFRVWREARDRIYVGPGRYHAMD 780
Db 721 EKNSLNRRFLPWNIEETEAISIDDDAHLRDEIMFGFRVWREARDRIYVGPGRYHAMD 780
QY 781 PHOSWLVNSYSCELSVNLGAAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFV 840
Db 781 PHOSWLVNSYSCELSVNLGAAFFHKYAYLYSYVMPQAIRDMVDEYINCEDIAMNFV 840
QY 841 SHITRKPIKVTSMWTRCPCGCPQALSHDSDSHHERHKCINFEVYKYGWMLLYTOFRVD 900
Db 841 SHITRKPIKVTSMWTRCPCGCPQALSHDSDSHHERHKCINFEVYKYGWMLLYTOFRVD 900

```

RESULT 2
 Q9WVL6 PRELIMINARY; PRT: 918 AA.
 AC Q9WVL6;
 DT 01-NOV-1999 (T-EMBLrel. 12, Created)
 DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
 DE EXOSTOSIN EXTL3.
 GN Mus musculus (Mouse)
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBITaxID:10990;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Sato T.;

Qy 901 SVLEKTRLPDKTKCKEFTI 919
 Db 900 SVLEKTRLPDKTKCKEFTI 918

RESULT 3
 ID Q9X208 PRELIMINARY; PRT: 972 AA.
 AC Q9X208;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE BCDNA:LD21192 PROTEIN.
 GN BOTV OR BCDNA:LD21192 OR CG15110.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 OC Ephyroidae; Drosophilidae; Drosophila.
 NCBI_TextId=7227;

[1]
 SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;
 MEDLINE=20196006; Pubmed=10731132;

RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Norton J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Arroll J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Boulter J., Brokstein P., Brotler P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Donnes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertilla S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harits N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,
 RA Jatala M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kohler C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-T., Wasarman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).

[2]
 SEQUENCE FROM N.A.

RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,
 RA Abmayyan A., Arcalana T.T., Baxter E., Blazek R.G., Butenhoff C.,
 RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,
 RA Galle R.A., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,
 RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,
 RA Park S., Segueira A., Sethi H., Smit E., Svirskas R.R., Weinburg T.,
 RA Celinker S.E.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A8003797; AAF57601.1; -;
 DR EMBL: AF132161; AAD34749.1; -;

DR FlyBase: FBgn0027535; botv.
 DR Interpro: IPR004263; Exostolin.
 DR Pfam: PF03016; Exostolin; 1.
 DR Hypothetical protein.
 SO SEQUENCE 972 AA; 110999 MW; 2A111BD6912FFCCA CRC64;

Query Match 47.5%; Score 2314; DB 5; Length 972;
 Best local Similarity 48.1%; Pred. No. 3.5e-177;
 Matches 473; Conservative 153; Mismatches 252; Indels 106; Gaps 21;

Qy 11 GAGGCGGTC-----MLRKSNNRIRLTW-----LSFLVFLVTFEPLLAHYLTLD 55
 Db 20 GSGGGENACAPNSSSAQIRISMGEFRISMROFRYKLPWYLLMLFLVSCLAIRILEVQ 79
 Qy 56 EAD-----ENAKRIFGPVGNELCEVYKVDLCLIRISVSEELLQLEAKROELN 104
 Db 80 DAPPLDLHRSSPLLDAYEDFSAMRAGDLKRIE---EMVRIKTSVSYELNELSESRQKLO 136
 Qy 105 SETAKNLKTIACKKSLENAKODLLQKNVISQTEHSYKELMAONOKLSLPIRLPEKD 164
 Db 137 SDISQYNQKIEELKQELRLREQTLELRKISVEQAOVAREAVORNTPDLLPRSLPNT- 195
 Qy 165 DAGLPPEK-----ATRGCLHNCFDYSRCPLTSGFPVYVYDSQGFV---GSYLDPLVK 215
 Db 196 ---LPRKSNPTGMAASCEHNCFNHSCSLSGFPVYLDPEHSVORKYDIDGFLK 252
 Qy 216 QAFQATARAVYVYENADIACLYVLGE--MQEPVLR----- 252
 Db 253 TTKLQKQTVGNAHYIKDKPKACIYLVGELLLEDLLRNRYAQAQEAHQOSTPTLEND 312
 Qy 253 -PALEKQQLSLPWRKDGNNHYINLSRK---SDTONLYNVTSGAAMVOSTFTYVQY 308
 Db 313 CPVDMER-LVSLPYMGDGRNHYLLNLARDLTSRNPPLRYQMTMAIVQSAFEERQF 371
 Qy 309 RPFGLVSPVYVHMSDPNFEIPEPOVPRKRYLFTFQEGEKIESRLSLOARFSEEME 368
 Db 372 RFGIDLVPPLGPGGDWQOEKEMPARKRYLLTQGE-LRKQSSLPNLDAPFILEHL 430
 Qy 369 GDEPADYDRIATLAKVADSKLDQVLEFTG---KNQPKSLPTMALGGEREDREL 424
 Db 431 AD-----MAKATDQGFVLQFCVATPQOEQSDSLP-DWTLGSGSSSRROL 475
 Qy 425 LKSLTFLIITPGDPRIVYSSGCTRFLEALGVAVPVUG-EDVOYLPYQDMLQMNAL 483
 Db 476 LKQSTFLIITPLNGR-VSSFLMARIVYELARSAVAVLIGADELRPLRYETVDMRTAL 534
 Qy 484 VVPRVYTEVHFLRSLSDSLAMRROGRFLTFEYFSTADSIENYVLAMRTPIQPA 543
 Db 535 LTPARITELHFLRAVQDADLLRLKROGLIMERTYSSVOATYDIYASLRDLGTPPR 594
 Qy 544 PIREAAAE-----IPRSGKAAGTPNMADNGDLGAPVETBPYPASPRYLNFTLV 597
 Db 595 PVPEVINGVSNVNFILKSDPPYGLDTEPEES---LGPI--PPVPSPAFRNRYTLR 648
 Qy 598 TDFRSMNACAPRPLHPHPEDVLPSEAKFLCGSGFRPIGAGGAGSGKEFQAALGN 657
 Db 649 MQAKEANNMDLPPYLYLPOLPFDLPALPSEAKFMGSHGFRIGLGLGAGKEFESESGN 708
 Qy 658 VPRQFTVVMVLYREVIVNLSLERLNGLPYLVNVMVNSPKLPSSDLMPDVGIVPM 717
 Db 709 YPRQFTVVMVLYREVIVNLSLERLNGLPYLVNVMVNSPKLPDLDLMPDVGIVPAV 768
 Qy 718 VRTKSNLNNFLPWNIEETALISIDDAHLRDEIMFGFRVREARDAIVGPGRYHA 777
 Db 769 LPARNRLNRRFLPFDVIEFPAVSVDDAHLRDELDFGRVVRERDRVYRPGRYHA 828
 Qy 778 WDI--PHQSWLYNSNYSCELSMVLGAAFPKHYAYLYSVYPOALIDMVDYENCIDIA 835
 Db 829 WDLGNPNQGMHYNSNYSCELSMVLGAAFPKHYAYLYLYTTLPPAIDKDYWNENCIDIA 888
 Qy 836 MNFLVSHITRPRIKYVSRMTFRGCGPOLSHDSHFHNHKKINFGVYVGYGMPLTY 895

Db 889 MNFLVSHTRKRPVKYTSRMTFRCPGCVSLSEDDTHFQERHAKCINFSTRVFGTYPLLNT 948
QY 896 QFRVDSVLEKTRLPDHTKCKEFTI 919
Db 949 QYRADSIIFKTRIPDHKKCKEFTI 972
RESULT 4
ID 017920 PRELIMINARY; PRT: 814 AA.
AC 017920;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE K01G5.6 PROTEIN.
GN K01G5.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RA Walli M.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL: Z92803; CAB07245.1; .
DR InterPro: IPR004263; Exostoin; 1.
DR Pfam: PF03016; Exostoin; 1.
SQ SEQUENCE 814 AA; 94196 MW; 38FA58C5EB17DB16 CRC64;

Query Match 25.7%; Score 1252.5; DB 5; Length 814;
Best Local Similarity 33.4%; Pred. No. 7,8e-92;
Matches 313; Conservative 147; Mismatches 312; Indels 165; Gaps 25;

QY 21 LRMSNRIRLTMISFTLVLIVFPPL--IAHYLT-----TIDEADEACKRI 64
Db 5 LNSSNSFVSLNVAFLRIFVITYITLYNNVSFSPSITODALKONTENDVDVASC 64
QY 65 FGRVGNELCEVKNHVDLCRIRESVE--ELLQLEAKROELNSETAKLNKTEACKSIEN 123
Db 65 SGYSIGRIILREOKRIILASVRLTESQVKIEIRTVQELQRLIPKQLESLAEGIEA 124
QY 124 AKODLLQKNVISQTHSYKELMAONPKLSLPIRLPEKDDAGLPPPKATRGCRHNCF 183
Db 125 AKQQL-----ELKRETONVKVFLPS--PLQIPRELQPSQISPNQDDIIT 168
QY 184 DYSRCPLTSGFPVYVYDSDQFVGSYLDPLVKAQFATARANYVTENADIACIYLVG 243
Db 169 DYSRCSISFMPYV-----FOSKSIHVQASKI--RSR 259
QY 244 EMGEPPVLRPALEKQLYSLPHMKRTGHNHVIITLSKSDTONLLYNVSTGRAMVAOSTF 303
Db 184 ----DIIISGQSEKEMLVN-----FOEVIPNLVETPDKACIKIHISNGIAS--ENTTF 230
QY 304 YTVQVYRPGDLYVSPLVHAMSEPNFMELPPQVYVKKRYLTFPGEGKIESLRSSLOEARSF 363
Db 231 NSLFRVNG-----SPLITN-----FOSKSIHVQASKI--RSR 259
QY 364 EEMEGSDPPADYD---DRI--IATLKAVDSKLDQVLEFTCKN-----QPKPS--LPP 410
Db 260 -----DFPVDVNHIAVEKVDLPLPLPQRENLISLIVNTLENSAFSSLSAEPSRRPI 313
QY 411 ENALGGER---EDRLLEKLTSTFALITTPGPRVLISSGCAITRLPEALEVGAVPVVLGE 466
Db 314 VLVKSGQENKSLERRRRLIGSSFTFCLP-----SMFPODFLSSIQLOGCIPIITLSN 365

QY 467 QVQLPYQDMLQWNEALVYVPPRYTEVHLLRSLSDSDDLAMRROGRFLMETESTADSI 526
Db 366 SOLPLFDOLIDMRATYRLPLARLPEAHFVQSFEISDIIEMRRYGRGLFETYLADRL 425
QY 527 FNTVLAMIRTIQIIPAALIREBAALRPHRSGKA--AGTDPNMADNGDLD--LGPVTEPP 583
Db 426 ARSLTALRYKLOIPTREVRNQALPLFNSSTFAPKSGSVVVOANFDEYLLGPLESR-- 483
QY 584 YASPRYLFNLTLYTDFRSWNCAPGPHLPNHPDFDLVSEAKFL--GSGTGRPIGGG 642
Db 484 VESTSYANFTFEPOLYSTDFNNIIMSPTYTEFLVNAALTELEFEDTIGFRPI--- 540
QY 643 AGSGGKEFOALGQNVPREQTYVMTLYEREVLNLSERLNGPLVYLNKVVVNSKPLP 702
Db 541 EPGSGAERSKALGQNRQEQFTVLLTYERDAVLGALERJHQPLINKIIVNNNNVRD 600
QY 703 SEDLLPDPICPVIVVTRKNSLNNRPLPMWETEALISDDAHLRHDEIMGFYWR 762
Db 601 PPD--TWPSLHIPVEIRVAENNLNNRVPMDRIETEAVALSDDDIDLMQOEIIIAFRWR 659
QY 763 EARDIVGEPGRYHAMDIPHQSMLYNSVSCSLMVLGAFFKRYAYLYSYVMPQAIR 822
Db 660 ENRRIYGFARHHA--RYGDSMFYNSNHTCOMMILTGAALFIKNTLYATYEMPAIR 717
QY 823 DMVDEYNCEDIANMFLVSHTRRPPIKVTYSRMTFRCPGCGQALSHDSHFERRHCKINF 882
Db 718 EHVNSIKNCEDIANMFLVSHTRRPPIKVTYSRMTFRCPGCGQALSHDSHFERRHCKINF 777
QY 883 FVKYGYTMPLLYQFRVDSVLEKTRLPDHTKCKEFTI 919
Db 778 FTKIYGYTMPLKFSQFRADSIIFKTRIPDHKKCKEFTI 972
RESULT 5
ID 0923D6 PRELIMINARY; PRT: 718 AA.
AC 0923D6;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE EXOSTOSES (MULTIPLE) 2.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RA SEQUENCE FROM N.A.
RP Strausberg R.;
RL Submitted (Aer-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC006597; AA06597.1; .
SQ SEQUENCE 718 AA; 82034 MW; 58722A09BD159CFB CRC64;

Query Match 20.6%; Score 1003; DB 11; Length 718;
Best Local Similarity 32.2%; Pred. No. 7,8e-72;
Matches 260; Conservative 125; Mismatches 255; Indels 168; Gaps 27;

QY 133 NVISQTHSYKELMAONPKLSLPI--RLPEKDDAGLPPPKATRGCRHNCFPYRCPILT 191
Db 49 SISSSDGVEKRSIREVPVYRLPTSPIDRGL-----SCRMHCFEYVYC--- 96
QY 192 SCF-----PVIYVSDQVY--FGSYLDPLVKAQFATARA---NVYTEMADIACIYVI 240
Db 97 -GPNPKNIKIVYIYPLKRYVDAGVYSSAISREYNELLTALISDSYTYDDINRACLPV 155
QY 241 LVGEM--QEPVYLRPALEKQLYSLPHMKRTGHNHVIITLSKSDTONLLYNVSTGRAMV 298
Db 156 SIDVNLQNP--LRIKETQAQLSLRW--DRGTNHLFNMPLGAPDPYNTALDIPRDRALL 212
QY 299 AOSTFYVYRPGFDL---VSPVLYHAMSEPNFMELPPQVYVKKRYLTFPGEGKIESLRSS 355
Db 213 AGGGSFTWYRQGVDSIVFSPLSAEMA-----LPERAPGPRRYFL-----LSS 257

Db 720 SQMRDLPLVLEKDOY 733

RESULT 9

ID 09JK82 PRELIMINARY; PRT: 746 AA.

AC 09JK82;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE EXOSTOSIN 1.

OS EXST1.

OS Cricetus griseus (Chinese hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;

OC Cricetus.

NCBI_TaxID=10029;

NCBI [1]

SEQUENCE FROM N.A.

Wei G., Bai X., Esko J.D.,

"Expression Cloning of Heparan Sulfate Copolymerase (EXT1) and Analysis of Mutations in Chinese Hamster Ovary Cell Mutants Defective in Heparan Sulfate Biosynthesis."

RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF52858; AAF7126.1; -

DR InterPro: IPR004263; Exostosin.

DR Pfam: PFO3016; Exostosin; 1.

SQ SEQUENCE 746 AA; 86188 MW; C3697BA4A21DA4F2 CRC64;

Query Match 17.98; Score 870.5; DB 11; Length 746;

Best Local Similarity 28.88; Pired. No. 3.9e-61;

Matches 233; Conservative 141; Mismatches 231; Indels 201; Gaps 31;

Db 162 EKDDAGL-PPPKATRG-----CRINHCDFYSCPLTSGFPVYV--DSQPFVGS 208

Db 72 ENESGHNVSPPKRDANSSYYKCKKCMSCFDFALCK-KNGKVVYVPOQKGEKTAES 130

Db 209 YLDELVQAFOATARAVYVETENADIAVLVGEVDEPVLRPA--ELEKOLYSLPH 265

Db 131 Y-----QNTLAIEGSEFTYSDPSOACFLVLSLDTLDRD-QLSPOVYVHNRSKVSQHL 183

Db 266 WRTGSHNVITNL--SRKSTQVNLVNSTGRMAVASTTYVQYRPGFPLVSPV-- 320

Db 184 W-NNGRNHLVFNLSYGTWPDYEDVDI--GQAMLAKASISTENFRNPFVSI-PLFSK 239

Db 321 ---HAMSEPMFEIPQVPRVRYKLYFTFOGEKI-----ESLRSSLOEARSEEGEDPP 372

Db 240 DHPRTGGERGLK-NTIPPLAKMLVFKKRYLTGIGSDPRMLVHVNEDVL----- 293

Db 373 ADYDDRIATLK-----AVODSKLDQVLEFETCKNPKPSLPTEMALCGEREDRLK 426

Db 294 -----LTTCKHGCKDMQKHKDSRCDRNTX-----EKYDREMLH 329

Db 427 LSTFALLITPDDPRVLVSSGATRLFALEVGAVPVVLGBOVLPYODMLDMNDAALVVP 486

Db 330 NATCLV-PPGRRL-----GSFRFLALAAVPMVLSNCGMELPFESEVIMWNOAAVIGD 382

Db 487 KPRTEVHFLRLSDSLDLMARQGRFLMETYFSTADSFNTVLAMIRTIQIPAPIR 546

Db 383 ERLLOIPSTRISHOKILALROTOFLMEAYFTSSVEKYVLTLEIIONDI----- 434

Db 547 EEAALIEPHRSKAAGTDPNMADGDLGLGVETEPVYASPRVLRNFTLVTVDFEYSGNC 606

Db 435 -----FKHIS-----RNSLI-----WNK 447

Db 607 APGFPHLPFPDPVLPSEAKPIG-----SGTGFRPIGGAGSGSGEPALAGVNP 660

Db 448 HPG--GLF-----VLPQYSSYLGDPEYYAANGLRP-----PSKFAVAVIHAVTP- 489

Db 661 EQFVVMVLTREVLVMSLERLNGLPYLNKVVVWNSPK-LPSEDLIMPDIQVPIVVR 719

Db 490 -----LVQSQPVYKLVLAAKAASQYCAQIYVLMNCDKRLPAKH-RWPAIVAVYIE 540

Db 720 TEKNSLNRRLEPMNEIETAILSDIDDAHLRHDEIMGFVRVREARDRIYGFPGRYHMD 779

Db 541 GESKVMSSRFLEPYNITTDVAVLSDIEDTVLSTTEVDFAFYVWSFPRYIGYPARSHFMD 600

Db 780 IPHOSMLYNSYSCSELSMVLTGAAFFHKYAYLYSYVPOAIRDMDEYINCEDIAMNFL 839

Db 601 NSKERMGYTSKMTNDYSVMVLGAIAVHKYHYLYTHYLPASLKNMVQDLNCEIDILNLF 660

Db 840 VSHTRKPKPKVT-----SNWTRPCGCGQALSHDSSHFERHKCINFP 883

Db 661 VSAVTKLPPIKVTOKKQYKETMNGQTSRASMA-----DPDHRQRQSCMNTF 708

Db 884 VKVGYMPLLYTOFRVDSVLEKTRL 908

Db 709 ASWEGYMLPLHSQMRDLPLVLEKDOY 733

RESULT 10

ID 09Y169 PRELIMINARY; PRT: 717 AA.

AC 09Y169;

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE BCDNA:GH02288 PROTEIN.

GN BCDNA:GH02288 OR CG8433.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydrioidae; Drosophilidae; Drosophila.

NCBI_TaxID=7227;

NCBI [1]

SEQUENCE FROM N.A.

RP STRAIN=BERKELEY.

RC MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blaziel R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Aroll J.F., Agbayani A., An H.-U., Andrews-Plannkoch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borokova D., Botchan M.R., Bouck J., Brokslein P., Brotlier P.,

RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Padios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Honck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iregyan C.,

RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Markilov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen T.,

RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,

RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Stryckas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,

RA Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.N., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,
RA Agabayan A., Arcaina T.T., Baxter E., Blazek R.G., Butenhoff C.,
RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,
RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Pacleb J.M.,
RA Park S., Sequeleira A., Sethi H., Snir E., Svirskas R.R., Weinburg T.,
RA Celniker S.E.;
RT "Full length Drosophila melanogaster cDNA sequence."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF003808; AAF58056.1; -
DR EMBL: AF145598; AAD38573.1; -
DR FlyBase: FBgn0027613; BCNDA:GH02288.
DR InterPro: IPR004263; Exostosin.
DR Pfam: PF03016; Exostosin; 1.
SQ SEQUENCE 717 AA; 82725 MW; A3894493C16EF7DC CRC64;

Query Match 17.3%; Score 841; DB 5; Length 717;
Best Local Similarity 29.3%; Pred. No. 8.6e-59;
Matches 227; Conservative 126; Mismatches 238; Indels 184; Gaps 24;

QY 177 CRLNCFDYSRCPITSGEPVYVSDQFV-----FGSYLDPLVKAQAFQATARAANYVTE 230
DB 92 CTFWDCLNTKCE-HDRLKYIYPLQEFVDEQSDKATATLSSEYFQILEAVLKS-R-YTTS 149
QY 231 NADYACLVILVGEVMOEPVVLRAPELEKO-----LYSLPHRTGHHNVIIN-LSRKS-DT 284
DB 150 NPNACFLPISLIDLNNV-----FDKHLAAGALASIDFW-DGSAHHIITNMLPGCAPS 202
QY 285 QNLLYVSTGRANVAOSTFTYVOYRPGFDLVY---SPLV---HANSEPNFMEIRPOVP 336
DB 203 YNVLVDVNTDAILFFGGFDSWSYRPGFDVAIPWSPRLVQAHAAQAQRKFLVVAQLN 262
QY 337 YKRYLTFPGGEKIESLRSIQEAKSFEEEMEGDPADYDRIITATLKAQODSKLDQVLY 396
DB 263 IL-----PRFVRTRELSLAHSEQLL 284
QY 397 EFTCKNPKPSLPTEMALGGERDRLE--LLKLSFALI---TTPDRLPVISGCATR 450
DB 285 LGACENIDL---TMRCPISQHHKSLSEYPRLSRGKFCILGRSLRMQOPDLY----- 332
QY 451 LFEALGAVNPVVLGEOVLPYODMLQWNEALVVPKRYTEVHLLRSISDSDDLAMRR 510
DB 333 --ELMSQHCIPVIAVDNYVLPFEDVIDMSLAVRIRENEHLSVMQKLAISVXIVEMOK 390
QY 511 QGRFLMETYESTADSIINTVLAIRIQIAPAIREEAAAEIHRSGKAAGTDPNNADN 570
DB 391 QVQWLEFSKYRKDKTLYLTALVELESRI---PPLRARS----- 426
QY 571 GDLDLGEVTEPPYASPRYLRNFLLTVDFYRSWNCAPGFHLPHPNPPDV-LPSEAKF 629
DB 427 -----RQMN---IIDNARSTFNPFLPSLA-- 449
QY 630 LSGGTGRPIIGGAGSGKEFOAALGQNVPRQ-FTVVMLTREEREVLNMSLERLNGLPY 688
DB 450 -----PKSGQFVAILTYDVRESLFLILQIAVPS 480
QY 689 LNKVVVVVNS-PKLPSDDLMPDIGVPIVYVTEKNSLNRLPWNEIETFAIISIDD- 746
DB 481 LOSILVIMNNOKSPHLLSTFPSSIKPLKIRQTKENKLSNRFYYPYPIETFEALITDDDI 540
QY 747 AHRHDEIMNGFWRREARDAIVGPGGRYHAMDIHQSMVLYNSVYSELNMTLGAFFH 806
DB 541 IMTLTDELDFGYEWMREPDPIHVGFSRIHWENVTMRMHESEMTQOISMVLTGAFFH 600
QY 807 KYVAYLYSVYVQAIRDMVDEYINCEDIANFLVSHTRKPIKVTSRMTFFRPGC--PQ 864
DB 601 KYVSHYTHAMPQDIDKQWDEHNCEDIANFLVANTNNPPIKVTPRKKFKKCECINTE 660
QY 865 ALSHDSHFHERHKCINFEVAYGYMPDLYTQFRVDSVLEFKTRLPHDKTKCFKFI 919

DB 661 MLSADLNMHREBSACIDRFESKTYGRMPLRTVEFRADPVLFRDNFP-DKIKRYNDI 714

RESULT 11

ID 098SV5 PRELIMINARY; PRT; 738 AA.

DT 01-JUN-2001 (TREMblrel, 17, Created)
DT 01-JUN-2001 (TREMblrel, 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel, 19, Last annotation update)

DE EXOSTOSIN.

OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.

NCBI_TaxId=8355;

RN [1]

RP SEQUENCE FROM N.A.
RA Hill A., Brown N., Hill M., Wells D.E.;
RT "Identification of a new ext family member in Xenopus laevis."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF319538; AAK15278.1; -
DR InterPro: IPR004263; Exostosin.
DR Pfam: PF03016; Exostosin; 1.
SQ SEQUENCE 738 AA; 84955 MW; D6F68E65720D085 CRC64;

Query Match 17.1%; Score 832; DB 13; Length 738;
Best Local Similarity 29.2%; Pred. No. 4.8e-58;
Matches 244; Conservative 123; Mismatches 236; Indels 232; Gaps 35;

QY 150 QPKLSLP-IRLLPEKDDAGLPP-----PKATRG-----CRLNCFDYSRCPITSG 193
DB 47 QPWLILPDLTSLPPTQEGGDPGQSVPRQKROSNSSITYGKRRMSCDLSIC-RRNG 105
QY 194 FPVYVYSDQFVGSYLDPLVKAQAFQATARAANYVTENADIACYVILVGEVMOEPVVL 253
DB 106 FKVYIYPOOK---GDKLSDSYOHLAIEISRFY-TSDPSQACVFLSLDTLDRD-OLSP 160
QY 254 ---AELKOLYSLPHMTDGHNVITNL---SRKSDQNLVYVSTGRANVAOSTFTYVQ 307
DB 161 HYVHNLSKVQONLHLN-NGGRNHLIFNLVSGTWPDTYEDVFDI--GQAMIAKASISTEN 217
QY 308 YRPGFDLVSPVYHAMSEP-----NFMETPPQVYKRYLTFPGGEKI-----ESTL 353
DB 218 FRPWFQD-VSIPLEKSHPRTGDKGFLRFNNIPM---RKVLYVKKGRYLTIGISDT 271
QY 354 RSSIQEAKSFEEEMEGDPADYDRIITATLKAQODSKLDQVLYEFTCKNPKPSLPTFW- 412
DB 272 RNALYHVIMGED-----VVLLTCTCKHGKK-----WQ 297
QY 413 ---ALCG-----EREDRLLEKLSTFALITPGDRLVYSSGCAIRLEALEVAVPV 462
DB 298 KHKDARCDKQNAEYKIDYEMLNHATFCLV--PRGRL-----GSRFLLEALQACVPV 350
QY 463 VLGEVOVLPYODMLQWNEALVVPKRYTEVHLLRSISDSDDLAMRQGRFLMETYFST 522
DB 351 MLSNGWELPSEVYIDWQAAVIGDERLLQIPSTIRSIHQDKLALRQOQFLMEAVFSS 410
QY 523 ADSIEFNTVLAIRIQIAPAIREEAAAEIHRSGKAAGTDPNNADNGLDGLGVETEP 562
DB 411 VERKIVLTTEIIQDRI-----FKHIS----- 431
QY 583 PYASPRILRMFTLYVDYRSWNCAPGFHLPHPNPPDVLPSEAKFLG-----SGTGF 636
DB 432 -----RNSLM-----WKKHKG--GLF-----VLPQYSTLGGFPYVYSLTGA 466
QY 637 RPIGGAGSGKEFOA-----LGGNVPREQFTVVMLTREEREVLNMSLERLNGLPYIN 690
DB 467 KP-----PSKFTQCDPCPCNPPGYSVPTNPPAAACCTKSQ-----CWYIN 505
QY 691 KVVVVVNSPK-LPSEDDLMPDIGVPIVYVTEKNSLNRLPWNEIETFAIISIDDADHL 749

Db 506 --TVLMWCDPFLPAKH-RWPATTVPVIVIGESKVMSSRFLPDNDIGTDAVLSLEDETVL 562
 QY 750 RHDEIMGFVREARDRIYVPGFRYHAMDIPIHOSMLYNSYCELSMTVLTGAAPFKHY 809
 Db 563 STEVEDPAFTVWVSFPRIVGPARSHFMDSTERNQYTSKWTNDMSVLTGAIYHKY 622
 QY 810 AYLXSYVMPQALIDMVDYINCEDIANMFLVSHITRKPRIKVT-----S 853
 Db 623 HYLYTNVLPASLKNMVDLANCEDILMNFVSAVTKLPRIKYOKKOYKETMAGOFFRAS 682
 QY 854 RMFRRCGCCQALSHDSSHFRHKCNFVVKYGYVPLLYTOFRDVSLEFKRL 908
 Db 683 RMA-----DDHFAQROTCLNTFATWLGWMLPILLSQMRDLPVLEKQV 725

RESULT 12
 076796 PRELIMINARY: PRT: 760 AA.
 076796;
 01-NOV-1998 (Tremblrel. 08, Created)
 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE PUTATIVE TUMOR SUPPRESSOR HOMOLOG.
 GN TTV OR TOUT-VEU OR CG10117.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98328070; PubMed=9665133;
 RA Bellaiche Y., The I., Perrimon N.;
 RT "Tout-veu is a Drosophila homologue of the putative tumour suppressor
 RT EXT-1 and is needed for hh diffusion.";
 RL Nature 394:85-88(1998).
 DR EMBL: AF083889; AAC32397.1; .
 DR FlyBase: FBgn002045; tlv.
 DR InterPro: IPR004263; Exoscolin.
 DR Pfam: PF03016; Exoscolin; 1.
 SQ SEQUENCE 760 AA: 87037 MW: EA8C040B52B54DB CRC64;

Query Match 16.3%; Score 796.5; DB 5; Length 760;
 Best Local Similarity 27.0%; Pred. No. 3.6e-55;
 Matches 228; Conservative 129; Mismatches 284; Indels 205; Gaps 26;

QY 122 ENMK-----QDLIDKNIYSTHSEYKELMAQNP-K-LSLPIRLPEKDDAGLPPKATRG 176
 Db 40 ESAKDGVOYHEQLPSFLGANDHMOQLQLQSNQSKLSSKHLVTRKPD----- 88
 QY 177 CRHLNCFDYSRCLPTSGFPYVYVDSQDFVGYSLDPLVKAQFOATARANY----- 226
 Db 89 CRMETGCDPLRC--YDRFLVYIYPP-----EPLNSGGAAPPPTSANYQKILTAQES 137
 QY 227 -VYTEMNADICLVILVAGMEQEPVVLNP--AELEKOLYSLPHMRTDGHNVIIINLSKRS- 282
 Db 138 RYTSPTACTACFLVIGIDTLDRDLSLSEDYRVNPSRLARLPYV--NNGRNHIIINLXSGTW 196
 QY 283 -DTONLLINSTRAMVAOSTFTYVQRPFDLVNPLVH-----ANSEPNFME 330
 Db 197 PDVAENSLGFDAGEALILAKSMGVLLRHFQDYSI-PLFHKQPLPRAGATGVQSNMF-- 253
 QY 331 IPROVPYKRYLFTFOGEKI-----ESLRSSIQEARSFEEMGDPPADVDRIITL-- 383
 Db 254 -----PANKRYLAFKCKRIVHIGSETNSLFLHNGR-----DNLVLTTCR 286
 QY 384 -----KAVODSKLDQVLEVFETCKNPKSLPTEMALGGERDRLELKLSTFALITTPGD 438
 Db 297 HGKSMRELQDNCRDEDEBREX-----DAYDETLLQNSTFCLV--PRG 336
 QY 439 PRVYSSGCATRLPEALEVGAPEVVLGEVOVLFYODMLQNEALVVPKPRVTEVHFLLR 498

Db 337 RRL-----GSFRLEALQAGCIPVLLSNMAVLPFESKIDMKQAAIMADELLQLQVPIVR 391
 QY 499 SLSDSDLLAMRQGRFLMETYFSTADSIENFTVAMIRTRIQIAPAIIRBAAEIHRS 558
 Db 392 SIAPAEHIFALROQTQVLMERYFSGISIKYFTTTEILRER--LPDIYVRSSLV----- 441
 QY 559 KAAGTDPNMANCDLIDGPEVEEPYASPRYLNRNLTLYTDFYSNMCAPGFHLPP-HT 617
 Db 442 -----NNSSGALLLTPTRA 456
 QY 618 PEDPVLPEAKFLGSGTGFRPGGAGGSGKFEQALAGVNPREDQTVMLTIREBEVLM 677
 Db 457 DSSRYMP-----FLNSMGAEPRHNTYAVIYVQGAALGPAA-----ALY 496
 QY 678 NSLERLNGLPYLNKVVVWNSPK-LPSEDLMPDIDGVPINV-----RT 720
 Db 497 KLVRTITKSQFVERILVLMNADRPDLKRRMPTSHIPLHVISLGSTRSQAGPTSOIT 556
 QY 721 E-KNSLNNRFLPWNIEIETEAISIDDALHRLHDEIMGFVREARDRIYVPGFRYHAND 779
 Db 557 EGRPSISORFLPYDEIQTDAVLSLDEDAILNTDELDPATVWMDPFRIVGYPARAHFMD 616
 QY 780 IPIHOSMLYNSYCELSMTVLTGAAPFKHYAYLYSYVMPQALIDMVDYINCEDIANMFL 839
 Db 617 DSKNANGYSKMTNYSIVLTGAFYHRYNYLYTMWLSLLKTKVQOSSNCEDIIMNLL 676
 QY 840 VSHITRKPRPKVSRMTFRCPQALSHDSSHFRHKCNFVVKYGYVPLLYTOFRDVSLEFKRL 899
 Db 677 VSHVTRKPRPKVYQKRGLOGSGDSGFAMWDPHFIFOROSCLNTFAVVGIMPLIRSLNR 736

QY 900 DSVLEFR 905
 Db 737 DPMLYR 742

RESULT 13
 09V730 PRELIMINARY: PRT: 760 AA.
 ID 09V730;
 AC 09V730;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE TTV PROTEIN.
 GN TTV OR CG10117.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephyroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=BERKELEY;
 RC MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadiot E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck C.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Krevitz S., Kulp D., Lai Z.,
 RA Lasko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Pollard D.R., Pacleb J.M.,
 RA Palazuelo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Shue B.C., Siden-Klamos I., Saunders R.D.C., Scheeler F., Shen H.,
 RA Slater E., Spradling A.C., Stapleton M., Skupski M.P., Smith T.,
 RA Stryckas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-T., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.-N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2193(2000).
 DR EMBL: AE003814; AAF58236.1; -
 DR Flybase: FBgn0020245; tlv.
 DR Interpro: IPR004263; Exostosin.
 DR Pfam: PF03016; Exostosin: 1.
 SQ SEQUENCE 760 AA; 87308 MW; F6369FDBA206DEA9 CRC64;

Query Match

16.3%; Score 795.5; DB 5; Length 760;

Best Local Similarity 26.8%; Pred. No. 4,4e-55;

Matches 227; Conservative 132; Mismatches 282; Indels 205; Gaps 26;

QY 122 ENAK-----QDLQKNVISTEHSYKELMAONPK-LSLPRLPERKDDAGLPPKATRG 176
 DB 40 ESARDGVQHPEDLPSELGANDMOELQLONSKSLDSSKHLVTRKPD----- 88
 QY 177 CRHNCEYDSCRPLTSGEPYVYDSDQFVGSYLDPLVKAQFQATARAVY----- 226
 DB 89 CRMTCTCFTRC--YDFELYITTP-----EPLNSLGAAPPTSNQKILTAIOES 137
 QY 227 -YVTENADIACLVYLVGEMQEPVLRP--AELEKOLYSLPHWRTDGHVNIIMSKRS- 282
 DB 138 RYVTSDDPTACLFVLDITLDRDLSLSDYVRNVPSRLARLPYV-NGRNHITFLVSGTW 196
 QY 283 -DTONLLYNSTGRAMVASTFTYVQYRPGFDLVVSPLVH-----AMSEPNME 330
 DB 197 PDVAENSLGFDAGAILAKASMGVLQRHGFVSI-PLFHKQFPLRAGATGVOSNFF-- 253
 QY 331 IPPQVPRKRYLTFQGEKI-----ESLRSSLOEARSEFEEMEGDPADYDDRIATL-- 383
 DB 254 -----PANKKTYLAFKRRKRYVHGSETRNSLPHLNGR-----DMVLVTTCR 296
 QY 384 -----KAVQDSKLDQVLVEFTCKNQKPRSLPTEWALCGEREDLELLKSTFALIITPGD 438
 DB 297 HGKSMRELQNRCDDEDREY-----DRYDEFTLLQNSFCLV--PRG 336
 QY 439 PRVYISGGCATRLEALEVGAAPVVLGEOVLPYQDMLQNNALVYPRKRYVEVHLLR 498
 DB 337 RRL-----GSFRLEALQACIPVLLSNANVLPRESKIDMKQAAIWADERLLQVPIVR 391
 QY 499 LSLSDSLAMRROGRFLMETYFTSTADISFNTVLMIRTIRIQIPAPIREEAAAEIPIRSG 558
 DB 392 SIPIERFALRQOTQVLEMEYFSGIEKIVFTTEILIRER--LPDYPYRSSLY----- 441
 QY 559 KAAGTDPNMAKDGLGVETEPYASPRYLKRNFTLVTDYKRSWNCACGPFHLFP-HT 617
 DB 442 -----WNSSPGALLTLTPFA 456
 QY 618 PFDVLPSEAKFLGSGTGFPIGGAGSGKKEQALGAVNPRQFVWMLTYEREVLM 677
 DB 457 DSSRYMP-----FLNSKGAEPRHNTAVIYVQIGALGPNA-----ALY 496
 QY 678 NSIERLNGELPLKVVVYVNSPK-LPSEDLIMPDIQVPIWV-----RT 720
 DB 497 KIVRTITKSGFERIIVLMAADRPLPLKRWPRPISHIPLHIVISLGSTRSGAGPTSQTT 556

QY 721 E-KNSLNRRFLPMNTEITETAILSIDDAHLRHEIMGFRVWRBARDRIYFPGRYHAMD 779
 DB 557 EGRSISGRFLPYDEIQTDAVLSDDEDAILNTDELDFAYVWRDPEPRIVGYPARAHFMD 616
 QY 780 IPHOSWLYNSYSCSLSNVLTGAAPFHKYAYLYSYVPOAIRDMVDEYINCEDIAMNFI 839
 DB 617 DSKNAMGYSKWTYYSIVLTGAAFYHRYNYLYTNMLSTLLKTVQOSSNCEDILNML 676
 QY 840 VSHTRKPKIKVSRMWRFCGCPQALSHDSHERHKCINPFVXYGVWPLLYTOFRV 899
 DB 677 VSHTRKPKIKVTKRKQKRETRGSPNPDHFIOQSCINTFAAVFGIPIRLSRMLR 736
 QY 900 DSVLKF 905
 DB 737 DPMLYR 742
 RESULT 14
 ID 09JRV7 PRELIMINARY; PRT; 669 AA.
 AC 09JRV7;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE MULTIPLE EXOSTOSES-LIKE 1 PROTEIN.
 GN EXTL1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stickens D.J., Brown D., Evans G.A.;
 RT "Ext genes are differentially expressed in bone and cartilage during
 RT mouse embryogenesis,"
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF224461; AAF61913.1; -
 DR MGI:1888742; Extl1.
 SQ SEQUENCE 669 AA; 74008 MW; CD547AEF25B4EAB1 CRC64;

Query Match 13.1%; Score 639; DB 11; Length 669;
 Best Local Similarity 26.1%; Pred. No. 1,4e-42;
 Matches 203; Conservative 117; Mismatches 221; Indels 238; Gaps 29;

QY 171 PKATRG--CHLNHOFDSRCPISGFPYVYDSDQFVGSYLDPLVKAQFQATARAVYV 228
 DB 69 PRVSRGSGCTWGAACFPDSKCR-----GVLKTIYHSPAGPTSEAQRRI 111
 QY 229 TEN-----ADIACLVYLVGEMQ-----PVVLRPAELEKOLYSLPHWRTDGHNH 273
 DB 112 LDSLEGSRYALSAD-ACLLFLPSQDRACGCPPL-----PNW-NGGRNH 156
 QY 274 VIINLSKSDTONLLYNSTGRAMVASTFTYVQYRPGFDLVVSPLVHNSSEPNEMFIP 333
 DB 157 LVLSLVPAPCTR-----LQAMVVAEASPSDIFRGFIAL-----P 193
 QY 334 QVPRKRYLTFQGEKTESLRSSLOEARSEFEEMEGDPPA---DYDDRIATATKAVODSK 390
 DB 194 YLP-----EA-----HPLRGAPRKLOHSHQRPATILLAAVEK 227
 QY 391 -----LDQVLVEFTCKNQKPRSLPTEWALCGEREDLELLKSTFALIITPGDPL 441
 DB 228 GWRWRTSTHASACIMRCHQDQGP-----QQTYPGETLIPNATFCLL--PG---- 271
 QY 442 VISSGCATRLF-EALEVGAAPVVLGEOVLPYQDMLQNNALVYPRKRYVEVHLLRSL 500
 DB 272 ---HRSATSCFIALQACIPVLLSPRWELPFESEVIDMTAAIADERLPLQVLAALREM 328
 QY 501 SCSDDLAMRROGRFLMETYFTSTADISFNTVLMIRTIRIQIPAPIREEAAAEIPIRSGKA 560
 DB 329 LPSRVIALRQOTQVLEMEYFSGIEKIVHTTLEVIQDRIV-----G 368

QY 561 ACTDPNMADNGDLDGCVETEPYASPRYLNTLVTFDFRSMNCAPGP-HLEPHTPF 619
 Db 369 ASCHPSTLMW-----SPGAL-LALPTESTSLQD-----DFYHL----- 402
 QY 620 DVLPSAKFLGSGTGRPIGGAGSGKFEQAALGNNVREOFTVYMLYEREEVLMS 679
 Db 403 -----QGGG-----PGSSTAVIWWGASGESLKL 428
 QY 680 LERLNGLPYLNKYVWNSKPLPSEDLMPDVGPIWVYTERKNSLNRPNNIETEA 739
 Db 429 IQEYAGSRHQAQLILILNSKPPPP--RMEPIAVPLVIGHR-KYSNRRFPYSNISTNV 485
 QY 740 ILISDDAHLRHDEIMGFYVWREARDRIYGFGRYHAMDIPHOSWLYNSNYCELSWL 799
 Db 486 ILISDAQSTLSTSEVDFAFYVWQSFPERMVGLSGSHFDEAOGMGYRTGMTNDFSML 545
 QY 800 TGAFFHKYAYLYSVYVMPQAIQMDVEXINCIDIANFLVSHITRKPRPKYTSRWTFRC 859
 Db 546 TTAFAFYHYHTLFTSHLPALRTIADPTCVDLNPLVATVTKLPKVP--YGRH 603
 QY 860 PGC-----PQALSHDSHFHERKCIINEFVKYGYMPLLYTOFRVSVLFK 905
 Db 604 PEAVPMDSGPRVPEPQPLQD-----CINRLAAGFHMPLVSSQVRLDPVLFK 653

RESULT 15
 ID 09D7G5 PRELIMINARY; PRT; 131 AA.
 AC 09D7G5;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE ADULT MALE TONGUE CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY.
 DE CLONE:2310009010, FULL INSERT SEQUENCE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shingawa A., Shidota K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
 RA Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Felschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustlingich S., Hill D., Hornann W., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sesaki H., Sato K., Schoenbach C., Seta T., Shidota Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL; AK009257; BAB26174.1;
 SO SEQUENCE 131 AA; 14976 MW; E8ED3DB7BB40A7 CRC64;

Query Match 12.0%; Score 586; DB 11; Length 131;
 Best Local Similarity 96.6%; Pred. NO. 2e-39;
 Matches 112; Conservative 2; Mismatches 2; Indels 0; Caps 0;
 QY 1 MTGYTMLRNGAGCGGTCCLMRNSNRIRLTWLSFTLFIIVFFPLIAHYLTTLDEADEA 60
 ||||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||

Db 1 MTGYTMLRNGAGCGGTCCLMRNSNRIRLTWLSFTLFIIVFFPLIAHYLTTLDEADEA 60
 QY 61 GKRIFGPRVGNELCEYKHYLDLCRIRSVSEELLQLEAKROELNSEIAKLNTKIEA 116
 ||||||| :||||||||| ||||||||||| ||||||||||| ||||||||||| |||||||||||
 Db 61 GKRIFGPRAGSELCEYKHYLDLCRIRSVSEELLQLEAKROELNSEIAKLNTKIEA 116

Search completed: September 4, 2002, 15:02:40
 Job time: 299 sec

THIS PAGE BLANK (USPTO)



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 17:58:59 ; Search time 4303.49 Seconds
(without alignments)
19357.128 Million cell updates/sec

Title: US-09-809-920-3

Perfect score: 6172
Sequence: 1 GGGCGGTCCCTGAGCTTGGA.....ATTAATGGAGACTTTAACTC 6172

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

arched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database :

EST:*
1: em_estdb:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	904.8	14.7	964	9	AL520206 AL520206
2	893.4	14.5	928	9	AL520207 AL520207
c 3	870.6	14.1	919	9	AL522291 AL522291
4	844.6	13.7	871	9	AL522292 AL522292
5	810.4	13.1	874	9	AU131026 AU131026
6	764.8	12.4	803	9	AU125122 AU125122
7	736	11.9	769	10	B1554177 603235213
8	734.4	11.8	770	10	BM046274 603626067
9	725.8	11.9	743	9	AU140433 AU140433
10	715	11.6	763	10	B1196587 602755556
11	706.4	11.4	734	9	AU132259 AU132259
12	706	11.4	768	9	AU133145 AU133145
13	672.4	10.9	1000	10	BG115714 602317188
14	666.2	10.8	755	9	AU133571 AU133571
15	663	10.7	941	10	BF982138 602308950
16	662.4	10.7	750	10	BM050902 603633742
17	658.6	10.7	968	10	BG424375 BG424375

c 18	650	10.5	672	10	BE343959	BE343959 602015653
19	642.4	10.4	644	12	A0334883	A0334883 RPT1-11-3
20	637.2	10.3	885	10	BG423077	BG423077 602450281
21	636.4	10.3	775	9	AL528972	AL528972 AL528972
22	634.2	10.3	987	10	BG023953	BG023953 602303650
23	633.4	10.3	635	10	BG910960	BG910960 602808764
24	633	10.3	644	9	AW603974	AW603974 IL3-CT021
25	628.8	10.2	644	9	AW603972	AW603972 IL3-CT021
26	620.4	10.1	736	10	BG261277	BG261277 602373022
27	617	10.0	853	10	BF311835	BF311835 601897354
c 28	598.8	9.7	627	10	BE839880	BE839880 QVO-FN018
29	596	9.7	804	10	BI911162	BI911162 603067590
30	592.4	9.6	623	9	AW954438	AW954438 EST36508
31	589.2	9.5	694	10	BM047368	BM047368 603628522
32	588	9.5	588	10	BE879757	BE879757 601491661
c 33	581.4	9.4	658	12	AG165871	AG165871 Pan trogl
34	570	9.2	957	10	BE897823	BE897823 601438486
35	560.8	9.1	742	10	BE142805	BE142805 MRO-HY015
36	557	9.0	557	9	AL046695	AL046695 DFFZP434C
37	556	9.0	616	10	BE264316	BE264316 601190146
38	550.6	8.9	579	10	BF311163	BF311163 601898629
39	546	8.8	569	10	BE383505	BE383505 601298567
40	543	8.8	595	9	BE142805	BE142805 MRO-HY015
41	538.6	8.7	593	10	BE265859	BE265859 601193042
42	534	8.7	546	10	BE278903	BE278903 601156314
43	531.8	8.6	704	10	BG924082	BG924082 602823672
c 44	529.8	8.6	559	9	AU148856	AU148856 AU148856
45	529	8.6	529	10	BF725639	BF725639 bx17e11.y

ALIGNMENTS

RESULT 1
AL520206/c
LOCUS AL520206 964 bp mRNA linear EST 13-FEB-2001
DEFINITION AL520206 LTI_NFL004_NBC2 Homo sapiens CDNA clone CS0DB005YN16 3
prime, mRNA sequence.
ACCESSION AL520206
VERSION AL520206.1 GI:12783699
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 964)
AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES

source
1. 964
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB005YN16"
/clone.lib="LTI_NFL004_NBC2"
/sex="male"
/tissue="neuroblastoma cells"
/lab.host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email: fliang@lifetech.com URL : http://fulllength.invitrogen.com" 5 others
BASE COUNT 227 a 239 c 272 g 161 t

FEATURES	source
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1. (bases 1 to 919)
TITLE	Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
JOURNAL	Full-length cDNA libraries and normalization
COMMENT	unpublished (2001)
LOCUS	Genoscope - Centre National de Sequencage
VERSION	BP 191 91006 EVRY cedex - France
KEYWORDS	Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr.
SOURCE	Location/Qualifiers
ORGANISM	1. 919
REFERENCE	/organism="Homo sapiens"
AUTHORS	/db_xref="taxon:9606"
TITLE	/clone="CSODB008YN08"
JOURNAL	/clone_11b="LTI_NFL004_NBC2"
COMMENT	/sex="male"
LOCUS	/tissue.type="neuroblastoma cells"
VERSION	/lab_host="DH10B"
KEYWORDS	/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-D119(dT) primer. Five prime end
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
LOCUS	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	

[illegible]

|||||
Db 81 TGTGGGGGAGCTTTGTATCATGCCACGCCACCACCCCAATACGTMACGA 22
OY 6113 GCACATGCTGTACATTTGA 6133
|||||
Db 21 GCACATGACTTACATTTGA 1
|||||
RESULT 4
AL522292 871 bp mRNA linear EST 13-FEB-2001
LOCUS AL522292 LTI_NFL004_NBC2 Homo sapiens cDNA clone CSDB008YN08 5
DEFINITION Prime, mRNA sequence.
ACCESSION AL522292
VERSION AL522292.1 GI:12785785
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 871)
AUTHORS Li, B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segrife@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Source
1. 871
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDB008YN08"
/clone_1lb="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact: Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email: fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 204 a 200 c 246 g 220 t 1 others
ORIGIN

Query Match 13.7%; Score 844.6; DB 9; Length 871;
Best local Similarity 99.9%; Pred. No. 9,7e-195;
Matches 844; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGGGGGCTCTGAGCTGGAAGCGGAGAGCAAGCCCTGAGGTTCACTTTTCAAGAGT 60
|||||
Db 27 GGGGGGCTCTGAGCTGGAAGCGGAGAGCAAGCCCTGAGGTTCACTTTTCAAGAGT 86
|||||
OY 61 CGGTGCTGAGGTGTATGCTACACAAGTGAAGGAAGAGGCTCTGAAACATGCG 120
|||||
Db 87 CGGTGCTGAGGTGTATGCTACACAAGTGAAGGAAGAGGCTCTGAAACATGCG 146
|||||
OY 121 CTGATTGTTGGCAAGCATCATAGAGCTGCAATTTATTTCTGTTCACTATTACT 180
|||||
Db 147 CTGATTGTTGGCAAGCATCATAGAGCTGCAATTTATTTCTGTTCACTATTACT 206
|||||
OY 181 GATAAAGTGTGAATAGACATATTTGTTGTCACAAACCAAGAACAGAGC 240
|||||
Db 207 GTTAACTGTGAATAGACATATTTGTTGTCACAAACCAAGAACAGAGC 266
|||||
OY 241 TATGCAATTTGAAAAGTCTGTGATTCAGAGGCTTTTCTGGGTTTCATCATCAG 300
|||||

Db 267 TATGCAATTTGAAAAGTCTGTGATTCAGAGGCTTTTCTGGGTTTCATCATCAG 326
OY 301 TACCTCTCCCTTCATCTCAGCAAGAAATGTGGACCTTTTATGTTGATTAAGATTAA 360
|||||
Db 327 TACCTCTCCCTTCATCTCAGCAAGAAATGTGGACCTTTTATGTTGATTAAGATTAA 386
|||||
OY 361 GGAATGTTCTTTGTCACAGCAGCAAGAAATGCTGTAATAGGTCAGAGACA 420
|||||
Db 387 GGACATGTTCTTTGTCACAGCAGCAAGAAATGCTGTAATAGGTCAGAGACA 446
|||||
OY 421 TTTGAGTGGCAGTGGAGAAAATGAATGTCATTTATTTGTCCTGTCCTGGAGAC 480
|||||
Db 447 TTTGAGTGGCAGTGGAGAAAATGAATGTCATTTATTTGTCCTGTCCTGGAGAC 506
|||||
OY 481 ACACCTAATCTTTCTGGAACAGTGCAGTGAAGACAGATCGTTTGTGAATAGCAACC 540
|||||
Db 507 ACACCTAATCTTTCTGGAACAGTGCAGTGAAGACAGATCGTTTGTGAATAGCAACC 566
|||||
OY 541 ATGCTATGGCAGTGAACCCGACGCTGATCTGGGGGAGAGCTGCAGAGACTATGACAG 600
|||||
Db 567 ATGCTATGGCAGTGAACCCGACGCTGATCTGGGGGAGAGCTGCAGAGACTATGACAG 626
|||||
OY 601 GCTATACCATGCTCGGAATGAGGGGCGCGGGAACGAGATGCACTGCTGCGCT 660
|||||
Db 627 GCTATACCATGCTCGGAATGAGGGGCGCGGGAACGAGATGCACTGCTGCGCT 686
|||||
OY 661 GGTCACACCGCATCGGCTGCTACGCTGATCTGACGCTCTTGTATCTGCTGCTTCT 720
|||||
Db 687 GGTCACACCGCATCGGCTGCTACGCTGATCTGACGCTCTTGTATCTGCTGCTTCT 746
|||||
OY 721 TCCGCTCATCGCCCACTATTACTTACCTACCACTCTGTGATGAGGCTGATGAGGACGCAAGC 780
|||||
Db 747 TCCGCTCATCGCCCACTATTACTTACCTACCACTCTGTGATGAGGCTGATGAGGACGCAAGC 806
|||||
OY 781 GGATTTTGTGCTCCCGGGGTGGGAGACAGCTGTGCGAGGTGAAGCAGCTGCTGATCTGT 840
|||||
Db 807 GGATTTTGTGCTCCCGGGGTGGGAGACAGCTGTGCGAGGTGAAGCAGCTGCTGATCTGT 866
|||||
OY 841 GCCGC 845
|||||
Db 867 GCCGC 871
|||||

RESULT 5
LOCUS AU131026 874 bp mRNA linear EST 24-OCT-2000
DEFINITION AU131026 NT2RP3 Homo sapiens cDNA clone NT2RP3001873 5', mRNA
sequence.
ACCESSION AU131026
VERSION AU131026.1 GI:10991380
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 874)
AUTHORS Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
Yamamoto, J., Makanetsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
Isigai, T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
Location/Qualifiers

```
source 1..874
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3001873"
/cell_type="teratocarcinoma"
/clone.lib="NT2RP3"
/cell_line="NT2"
/Note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 176 a 221 c 219 g 254 t 4 others
ORIGIN
```

```
Query Match 13.1%; Score 810.4; DB 9; Length 874;
Best Local Similarity 97.3%; Pred. No. 2.1e-186;
Matches 854; Conservative 0; Mismatches 20; Indels 4; Gaps 3;
```

```
3691 TATTAAAAAGAAAGTTTCAATTTGCCATTCAGAGCTTATTATATATATGTGTG 3750
|||||
1 TATTAAAAAGAAAGTTTCAATTTGCCATTCAGAGCTTATTATATATATGTGTG 60

3751 TATATAATACATGACACACTTGATATATATTTTGGCTGGGAGTGTAGTT 3810
|||||
61 TATATAATACATGACACACTTGATATATATTTTGGCTGGGAGTGTAGTT 120

3811 TTGCTTTCTAAGGAGGAGGAGCGGAGGCTCTTTGTTCTGTATTTGCGGAGATGG 3870
|||||
121 TTGCTTTCTAAGGAGGAGGAGCGGAGGCTCTTTGTTCTGTATTTGCGGAGATGG 180

3871 TTGCTGCTTGTGCTACTGCTTATGCTTAAATATATATCCATCTCCCGAGGCGCAT 3930
|||||
181 TTGCTGCTTGTGCTACTGCTTATGCTTAAATATATATCCATCTCCCGAGGCGCAT 240

3931 CTGTGTGACGACACAGAAAGGATGATGCTTGGCTTGTGCGGCTGTGACAGGCTTC 3990
|||||
241 CTGTGTGACGACACAGAAAGGATGATGCTTGGCTTGTGCGGCTGTGACAGGCTTC 300

3991 TTGCTTACCTTTCTGTGTGCTACTGACGACATCTGTATCTCATTTCTCCAGTGAAT 4050
|||||
301 TTGCTTACCTTTCTGTGTGCTACTGACGACATCTGTATCTCATTTCTCCAGTGAAT 360

4051 CCTGAGGAGGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4110
|||||
361 CCTGAGGAGGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420

4111 TTGAGGATTTATCTTTAGGCGAGGCTTGCCTCCGTAATTCCTGCTGCCATTTCTC 4170
|||||
421 TTGAGGATTTATCTTTAGGCGAGGCTTGCCTCCGTAATTCCTGCTGCCATTTCTC 480

4171 TCTTGTGAGAGAGATGAGAGCAAGAGTGAAGAAATAGGGGCTGAAGAGCCCA 4230
|||||
481 TCTTGTGAGAGAGATGAGAGCAAGAGTGAAGAAATAGGGGCTGAAGAGCCCA 540

4231 CTGCGAGATGCTCTTCTTCTGCTCTTCTGTAACACACGTCCTTGGGCTCCAG 4290
|||||
541 CTGCGAGATGCTCTTCTTCTGCTCTTCTGTAACACACGTCCTTGGGCTCCAG 600

4291 GCGTTTCTGAAGTCTCTTCTGATTTGAGACAGAGATGAGAGGCTGACATCTGCTG 4350
|||||
601 GCGTTTCTGAAGTCTCTTCTGATTTGAGACAGAGATGAGAGGCTGACATCTGCTG 660

4351 TGTGCTGAAGTCTTGTGACAGTACGCTCTCTCCCTAGTGTAGAGAACGAGTGTCT 4410
|||||
661 TGTGCTGAAGTCTTGTGACAGTACGCTCTCTCCCTAGTGTAGAGAACGAGTGTCT 720

4411 TCGAGGAAACCCAGCGCTGCGGAGAGTTTATACAGCAAGGCGCTGCTTGGGATAT 4470
|||||
720 TCGAGGAAACCCAGCGCTGCGGAGAGTTTATACAGCAAGGCGCTGCTTGGGATAT 779

4471 TCGTGTGTAATTTACACTTCCCGCGCTCTGTCTGAGAGCCCATCTGTATCTGT 4530
|||||
780 TCGTGTGTAATTTACACTTCCCGCGCTCTGTCTGAGAGCCCATCTGTGTGAGAGCCCATCTGTG 836
```

```
QY 4531 GGTTTTGGACCCCTAATGTACAGCTTGGCTGTAGACT 4568
Db 837 GGTTTTGGACCCCTAATGTAGNACTTGGCTGTAGACT 874
```

```
RESULT 6
AUI25122 803 bp mRNA linear EST 23-OCT-2000
LOCUS AUI25122 NT2RM4 Homo sapiens cDNA clone NT2RM4001073 5', mRNA
DEFINITION sequence.
ACCESSION AUI25122
VERSION AUI25122.1 GI:10949838
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

```
REFERENCE 1 (bases 1 to 803)
Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Salto,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and Isogai,T.
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Salto,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S., Isogai,T.)
Unpublished (2000)
CONTACT: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
```

```
JOURNAL HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
FEATURES
location/Qualifiers
source 1..803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RM4001073"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/Note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal precursor cells"
```

```
BASE COUNT 177 a 194 c 196 g 233 t 3 others
ORIGIN
```

```
Query Match 12.4%; Score 764.8; DB 9; Length 803;
Best Local Similarity 99.1%; Pred. No. 2.7e-175;
Matches 788; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
```

```
QY 3666 TACATTCATACACATATTATGATTTTAAAAAGAAAGTTTACAGATTTCCATTTCA 3725
|||||
9 TACATTCATACACATATTATGATTTTAAAAAGAAAGTTTACAGATTTCCATTTCA 68

3726 AGGCTATTATATATATATGCTGTATATATATATATATATATATATATATATATATAT 3785
|||||
69 AGGCTATTATATATATATGCTGTATATATATATATATATATATATATATATATATAT 128

3786 ATTTTGGCTGGGAGTGTAGATTTTGGCTTCTAAGGAGGAGGACCGCAGAGCTCTT 3845
|||||
129 ATTTTGGCTGGGAGTGTAGATTTTGGCTTCTAAGGAGGAGGACCGCAGAGCTCTT 188

3846 TGTTCGTATTTCTGGGAGAGATGGTCTGCTGTGTCACAGGCTTATCTTAAAGAT 3905
|||||
189 TGTTCGTATTTCTGGGAGAGATGGTCTGCTGTGTCACAGGCTTATCTTAAAGAT 248

3906 CATCTCCATCTCCCGAGGCGCATCTGTGTGAGCAACGAGAAAGGATGAACCTTGGCC 3965
|||||
249 CATCTCCATCTCCCGAGGCGCATCTGTGTGAGCAACGAGAAAGGATGAACCTTGGCC 308
```

OY	3966	CCTTGGCGGCGCCGTGGACAGAGTGTCCTTTCCCTTAACCCCTTCGTTGCAGCAGCAACTCGT	4025
Dd	309	CTCTTGGCGGCGCCGTGGACAGAGTGTCCTTTCCCTTAACCCCTTCGTTGCAGCAGCAACTCGT	368
OY	4026	AACATCATTTCTTCTCCACAGTAGAATCCCCTGGAGAGCCCTGACCCTTGSTGGCGTTCACG	4085
Dd	369	AACATCATTTCTTCTCCACAGTAGAATCCCCTGGAGAGCCCTGACCCTTGSTGGCGTTCACG	428
OY	4086	TTCCTGCTGCTTGGGGCCAGCCGATTTTTTGAGGATTATCTTTAGGCCAGGCTTGCCGT	4145
Dd	429	TTCCTGCTGCTTGGGGCCAGCCGATTTTTTGAGGATTATCTTTAGGCCAGGCTTGCCGT	488
OY	4146	ACTTATCCCTGCTCTCCACATTTCTCTCTGTTTGAGAGAGATGAGGAAGCAAGAAGTGA	4205
Dd	489	ACTTATCCCTGCTCTCCACATTTCTCTCTGTTTGAGAGAGATGAGGAAGCAAGAAGTGA	548
OY	4206	GAAAGATATGAGGGGTGAAGACGCCACATCCAGATGGCTTTCTATCCTGCTCTTCTGT	4265
Dd	549	GAAAGATATGAGGGGTGAAGACGCCACATCCAGATGGCTTTCTATCCTGCTCTTCTGT	608
OY	4266	GAAACACAGTGTCTGTGGGGCCTCAGGCGTTTCTGAAGTGTCTTCTTGTGATTTGACAGG	4325
Dd	609	GAAACACAGTGTCTGTGGGGCCTCAGGCGTTTCTGAAGTGTCTTCTTGTGATTTGACAGG	668
OY	4326	AGATCAGCAGCGCTGCATCTGCTGTGCTGTAAGTGTGTTGCAGTGCACCTCTCCCTGC	4385
Dd	669	AGATCAGCAGCGCTGCATCTGCTGTGCTGTAAGTGTGTTGCAGTGCACCTCTCCCTGC	728
OY	4386	CTAGTGTGAGAC -AAGCCAGATGTCCTTGAGAGAACCCACCGGCTGG -CCGGGAAGTTT	4443
Dd	729	CTAGTGTGAGACAAAGCCAGATGTCCTTGAGAGAACCCACCGGCTGGCCGGAAGTTT	788
OY	4444	ACAGCAAGCGCCCTG 4458 	
Dd	789	ATAGCAAGCGCCCTG 803	
RESULT	7		
LOCUS	B1554177	769 bp mRNA linear EST 05-SEP-2001	
DEFINITION	6032352J.F1 NCI CGAP_Brn67 Homo sapiens cDNA clone IMAGE:5309642		
ACCESSION	B1554177	5'..mRNA sequence.	
VERSION	B1554177.1	GI:15441480	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 769) NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov Tissue Procurement: David N. Louis, M.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM1784 row: 1 column: 03 High quality sequence stop: 754.		

BASE COUNT 153 a 217 c 216 g 182 t 1 others

ORIGIN

/lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr.
 Average insert size 2.3 kb. Constructed by Life
 Technologies. Note: this is a NCI-CCAP Library."

Query Match	Best Local Similarity	11.9%	Score 736	DB 10	Length 769
Matches 761	Conservative 0	Mismatches 6	Indels 2	Gaps 2	
QY 2005	CCTACGAGACATGTCGACATGGAGAGAGGCGCCCTGTTGTGTGCCAAACCTCGTGTGA				2064
Db 1	CCATACGAGGACATGTCGACATGGAGAGAGGCGCCCTGTTGTGTGCCAAACCTCGTGTGA				60
QY 2065	CCGAGGTTATTTCTTCGTCGACAGAGGCTCTCCGATATGTACCTCTGGCTATATAGGCGGC				2124
Db 61	CCGAGGTTATTTCTTCGTCGACAGAGGCTCTCCGATATGTACCTCTGGCTATATAGGCGGC				120
QY 2125	AAGGCGCTTCTCTCTGGGAGACTTACTTCTCCACTGCTACAGATATTTT-TAATACCGTG				2183
Db 121	AAGGCGCTTCTCTCTGGGAGACTTACTTCTCCACTGCTACAGATATTTTATATATACCGTG				180
QY 2184	CTGGCTATATATAGCATGCGATCCAGATCCACGCCGCTCCATCCGGGAAGAGCGGCA				2243
Db 181	CTGGCTATATATAGCATGCGATCCAGATCCACGCCGCTCCATCCGGGAAGAGCGGCA				240
QY 2244	GCTGAGATCCCCACCGCTTCAGGCAAGGGGGCTGGAACTAGACCCCAACATGGCTGCAAC				2303
Db 241	GCTGAGATCCCCACCGCTTCAGGCAAGGGGGCTGGAACTAGACCCCAACATGGCTGCAAC				300
QY 2304	GGGAGACTTGACCTGGGGCCAGTGGAGAGAGGAGCGCCCTACGCTCACCCAGATACCTC				2363
Db 301	GGGAGACTTGACCTGGGGCCAGTGGAGAGAGGAGCGCCCTACGCTCACCCAGATACCTC				360
QY 2364	GCGAATTTGACTGACATGTCACATGACTTTTACCAGTGGAACTGTGCTCCAGGCGCT				2423
Db 361	GCGAATTTGACTGACATGTCACATGACTTTTACCAGTGGAACTGTGCTCCAGGCGCT				420
QY 2424	TTTCATCTTTTCCCCCAGCACTCCCTTTTGAACCTGTGTTCCCTCAGAGGCCAAATCTTG				2483
Db 421	TTTCATCTTTTCCCCCAGCACTCCCTTTTGAACCTGTGTTCCCTCAGAGGCCAAATCTTG				480
QY 2484	GGCTAGGAGACTGCGCTTTCGGCCCTATTGCTGTGGAGCTGGGGGTTCTGGCAAGAAATT				2543
Db 481	GGCTAGGAGAACTGGCTTTCGGCCCTATTGCTGTGGAGCTGGGGGTTCTGGCAAGAAATT				540
QY 2544	CAGGAGGCGCTTGGAGGCAATGTTCCCGAGAGCAAGTTACGGTGTGATTTGACTTAT				2603
Db 541	CAGGAGGCGCTTGGAGGCAATGTTCCCGAGAGCAAGTTACGGTGTGATTTGACTTAT				600
QY 2604	GAGCGGAGAGAAAGTCTTATGAACTCTTTAGAGAGGCTG-AAATGCGCTCCCTTACCTGAA				2662
Db 601	GAGCGGAGAGAAAGTCTTATGAACTCTTTAGAGAGGCTGAAATGCGCTCCCTTACCTGAA				660
QY 2663	CAAGGTCGTGGGGGTGGGAATTTCTCCCAAGCTGCATCAGAGGACCTTCTGTGGCGCTGA				2722
Db 661	CAAGGTCGTGGGGGTGGGAATTTCTCCCAAGCTGCATCAGAGGACCTTCTGTGGCGCTGA				720
QY 2723	CATTGGCGTTCCTCCATCATGATGGGTCCTGACTGAGAAGAAAGATTTTGAAC 2771				
Db 721	CATTGGCGTTCCTCCATCATGATGGGTCCTGACTGAGAAGAAAGATTTTGAAC 769				
RESULT 8					
BM046274	770 bp	mRNA	linear	EST 07-NOV-2001	
LOCUS	603626067E1	NIH_MGC_40	Homo sapiens cDNA clone IMAGE:5452432 5'		
DEFINITION	mRNA sequence.				
ACCESSION	BM046274				
VERSION	BM046274.1	GI:16775541			

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE NIH-MGC http://mgi.ncl.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: c9abs-r@mail.nih.gov
Tissue procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Plate: LNCM944 Row: f Column: 17
High quality sequence stop: 770.
Location/Qualifiers
1..770
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH-MGC_40"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pOT87; Site:1: XhoI; Site:2: EcoRI; CDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

BASE COUNT 146 a 206 c 193 g 225 t
ORIGIN

Query Match 11.9% Score 734.4; DB 10; Length 770;
Best Local Similarity 99.5%; Pred. No. 7e-168;
Matches 766; Conservative 0; Mismatches 1; Indels 3; Gaps 3;

OY 4565 GACTCCCCGAGCTTGGTATGCTAGAACAAATGAGAGGCTGATTTGCTGTAAAGCT 4624
1 GACTCCCCGAGCTTGGTATGCTAGAACAAATGAGAGGCTGATTTGCTGTAAAGCT 60

OY 4625 CACATCCAGCCTTGATCTAACGGGCAATTCACAAACCGAGATTACCACTTCCACTCCCT 4684
1 CACATCCAGCCTTGATCTAACGGGCAATTCACAAACCGAGATTACCACTTCCACTCCCT 120

OY 4685 GCTTAGAGTTGTCCTCCGCGGCGAAGCTGAATTTTGGTGGTCAAGGCTG 4744
1 GCTTAGAGTTGTCCTCCGCGGCGAAGCTGAATTTTGGTGGTCAAGGCTG 180

OY 4745 CAGTAGGGGAAGCTAGAGGGGTGAGTGGCATTTGTACAGGATTTAGCCCATGACGTG 4804
1 CAGTAGGGGAAGCTAGAGGGGTGAGTGGCATTTGTACAGGATTTAGCCCATGACGTG 240

OY 4805 TTCTTGAACCTTACTTTTGGAGTGGAGTTGACTCTGGAAGTTTTCAGCACTGAACA 4864
1 TTCTTGAACCTTACTTTTGGAGTGGAGTTGACTCTGGAAGTTTTCAGCACTGAACA 300

OY 4865 AAAGCTCAGGTTTGTCTGGATGACATGACATGCTTAAAGCCAGTTCCGTTCCCTAGAC 4924
1 AAAGCTCAGGTTTGTCTGGATGACATGACATGCTTAAAGCCAGTTCCGTTCCCTAGAC 360

OY 4925 TTGGCATCTGCTGCTTCTTATTTCTTGAATACGTTCTCTGACCTCCGTTACACAGT 4984
1 TTGGCATCTGCTGCTTCTTATTTCTTGAATACGTTCTCTGACCTCCGTTACACAGT 420

OY 4985 GGGTCTCTTCAAGTACTGTTTTTGAAGCTGGGCTCTTTGTGTAGCTCCACCCACCTGT 5044

Db 421 GGGTCTCTTCAAGTACTGTTTTTGAAGCTGGGCTCTTTGTGTAGCTCCACCCACCTGT 480

OY 5045 AGGCTAGCTCGGCTTAAAGGAATCTCTCCCATTTGGCAAAACCGACCGCGCCGCAAG 5104
1 AGGCTAGCTCGGCTTAAAGGAATCTCTCCCATTTGGCAAAACCGACCGCGCCGCAAG 539

OY 5105 GACTGTGTTCCAAAGTTTCCCGCCGCCAACCCAGCAACCTGAGTCCCTGCT 5164
1 GACTGTGTTCCAAAGTTTCCCGCCGCCAACCCAGCAACCTGAGTCCCTGCT 598

OY 5165 GAGCAGTGTGTTATGTTCCAGAGTGGGGGTGACAGCCCTTCTCAGAACTTCTA 5224
1 GAGCAGTGTGTTATGTTCCAGAGTGGGGGTGACAGCCCTTCTCAGAACTTCTA 658

OY 5225 GTTGCCCTTACTGACTGCTGACTGATTTCTTTTGAAGTACCTTCTCCGCGG 5284
1 GTTGCCCTTACTGACTGCTGACTGATTTCTTTTGAAGTACCTTCTCCGCGG 718

OY 5285 GAGCC-AAAGTGTGTGTGGCGCTATATGTGGCTGCTATTTCATCTG 5335
1 GAGCCAAAGTGTGTGTGGCGCTATATGTGGCTGCTATTTCATCTG 770

RESULT 9
AUI40433 743 bp mRNA linear EST 25-OCT-2000
LOCUS AUI40433 PLACE3 Homo sapiens cDNA clone PLACE3000047 5', mRNA
DEFINITION sequence.
ACCESSION AUI40433
VERSION AUI40433.1 GI:11001954
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 743)
Ota,T., Suzuki,Y., Saito,K., Ishii,S., Yamamoto,J., Sugiyama,T., Nishikawa,T., Nakamura,Y., Sugano,S., Masuho,Y., and Isigai,T.
Yamamoto,J., Sugiyama,T., Nishikawa,T., Nakamura,Y., Sugano,S., Masuho,Y., Isigai,T.)
Unpublished (2000)
CONTACT: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.
location/Qualifiers
1..743
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="PLACE3000047"
/clone_lib="PLACE3"
/tissue_type="placenta"
/note="Vector: pME18SFL3"

BASE COUNT 154 a 204 c 201 g 181 t 3 others
ORIGIN

Query Match 11.8% Score 725.8; DB 9; Length 743;
Best Local Similarity 99.2%; Pred. No. 8.6e-166;
Matches 738; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

OY 2125 AAGCCGCTTCTGTGGAGACTTACTTCTCAGTGTGACAGTATTTTAAATACCGTGC 2184
1 AAGCCGCTTCTGTGGAGACTTACTTCTCAGTGTGACAGTATTTTAAATACCGTGC 60

```

found through the I.M.A.C.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM1769 row: p column: 14
High quality sequence start: 4
High quality sequence stop: 761.
Location/Qualifiers
1. 763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4891261"
/clone_1b="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; CDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using Zap-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT
144 a 230 c 218 g 171 t
ORIGIN

```

|||||
Db 662 CTTTTCATCTTTTCCCA-AGTCCCTTGACCCCTGTGGC--TCAGAGGCAATTC 718
Oy 2481 TTGGGCTCAGGAGCTGGCTTTGGCCATTGTTGGTGGAGCTGG 2525
|||||
Db 719 TTGGGCTCAGGAGCTGGCTTTGGCCATTATAGTGTGGAGCTGG 763
RESULT 11
AUI32259 734 bp mRNA linear EST 24-OCT-2000
LOCUS AUI32259 NT2RP3 Homo sapiens cDNA clone NT2RP3004092 5', mRNA
DEFINITION
sequence.
ACCESSION AUI32259
VERSION AUI32259.1 GI:10992613
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 734)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Salto,K., Kawai,Y.,
Yamamoto,J., Nakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
LOCATION/Qualifiers
1. 734
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP3004092"
/clone_1b="NT2RP3"
/cell_type="teratocarcinoma"
/note="Vector: PME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
FEATURES
SOURCE
ORIGIN
BASE COUNT 173 a 193 c 197 g 168 t 3 others
Query Match 11.4% Score 706.4; DB 9; Length 734;
Best Local Similarity 99.2%; Pred. No. 4.6e-161;
Matches 729; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
Oy 3032 TTTTATGTGATGCCAGGCGCATCCGGAGATGATGATGAATACATCACTGTGAGA 3091
|||||
Db 1 TTTTATGTGATGCCAGGCGCATCCGGAGATGATGATGAATACATCACTGTGAGA 60
Oy 3092 CATTGCCATGAACTTCTGTCTCTCCACATACATCGGAAGCCCCCATCAAGGTGACTC 3151
|||||
Db 61 CATTGCCATGAACTTCTGTCTCTCCACATACATCGGAAGCCCCCATCAAGGTGACTC 120
Oy 3152 AGGTGATGATCCGATGCCAGATGCGCTCAGGCCCTGTCTCATGATGATGATCCCACTT 3211
|||||
Db 121 AGGTGATGATCCGATGCCAGATGCGCTCAGGCCCTGTCTCATGATGATGATCCCACTT 180
Oy 3212 CCACGAGGCGCAAGATGATCAACTTCTGTGTAAGGTGTAAGGCTAATGCCCTCTCT 3271
|||||
Db 181 CCACGAGGCGCAAGATGATCAACTTCTGTGTAAGGTGTAAGGCTAATGCCCTCTCT 240
Oy 3272 GTACGAGATTCAGGGGTGATCTGTGCTCTTCAAGACAGGCTGCCCATGACAGAC 3331
|||||

Db 241 GTACAGCAATTCAGGATGATCTGTGCTCTTCAGACACGCCCTGCCATGACAAAGAC 300
Oy 3332 CAAGTCTCAAGTTCAATCATTAGGGGACAGCCAGCGTCTGGGGAAGAGATGACAGAGG 3391
Db 301 CAAGTCTCAAGTTCAATCATTAGGGGACAGCCAGCGTCTGGGGAAGAGATGACAGAGG 360
Oy 3392 AGGAAGTGGCTCCCAAGGTTCTTCTAGGCATTGCAAGACCTTTGGGACATCTGCTGGG 3451
Db 361 AGGAAGTGGCTCCCAAGGTTCTTCTAGGCATTGCAAGACCTTTGGGACATCTGCTGGG 420
Oy 3452 TGGGCCAGAGCTCTGCTGGAAGGGGAGAGAGGAGTGGAAGGAACGCTGCCCTTA 3511
Db 421 TGGGCCAGAGCTCTGCTGGAAGGGGAGAGAGGAGTGGAAGGAACGCTGCCCTTA 480
Oy 3512 TCTTGAAGTCAAGCACACTGAGGCTGAGCCCTGGAGCGGAGTCCCGGGGTTCCCAAC 3571
Db 481 TCTTGAAGTCAAGCACACTGAGGCTGAGCCCTGGAGCGGAGTCCCGGGGTTCCCAAC 540
Oy 3572 AGGCACTGACTGATGATGATCTTACACTGAGACCTGCGGACTCTGCAAGTCACTACAC 3631
Db 541 AGGCACTGACTGATGATGATCTTACACTGAGGCTGCGGACTCTGCAAGTCACTACAC 600
Oy 3632 GTTCGTACGCGCCAGGAC-AGCTGTGCTGTGTTTACATTCAATACACTATTATGAT 3690
Db 601 GTTCGTACGCGCCAGGACAGCTGTGCTGTGTTTACATTCAATACACTATTATGAT 660
Oy 3691 TATTTAAAGAGAAAGTTTCAGATTGCAATTCAGGCTTATTTATATATGATGTGTG 3750
Db 661 TATTTAAAGAGAAAGTTTCAGATTGCAATTCAGGCTTATTTATATATGATGTGTG 719
Oy 3751 TATTTAAATACATGC 3765
Db 720 NATTTAAATACATGC 734
RESULT 12
AUI33145 768 bp mRNA linear EST 24-OCT-2000
LOCUS AUI33145 NT2RP4 Homo sapiens cDNA clone NT2RP4001395 5', mRNA
DEFINITION
sequence.
ACCESSION AUI33145
VERSION AUI33145.1 GI:10993684
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 768)
Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Salto,K., Yamamoto,J.,
Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuko,Y. and
Isogai,T.
TITLE HRI human cDNA project (Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y.,
Salto,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano
S., Masuko,Y., Isogai,T.)
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
LOCATION/Qualifiers
1. 768
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2RP4001395"
/clone_1b="NT2RP4"
/cell_type="teratocarcinoma"

/cell_line="NT2"
/note="Vector: pME18SFL3; mRNA from NT2 neuronal precursor
cells after 2-weeks retinoic acid (RA) induction"
BASE COUNT 163 a 200 c 202 g 199 t 4 others
ORIGIN

Query Match 11.4% Score 706; DB 9; Length 768;
Best Local Similarity 97.9%; Pred. No. 5.9e-161;
Matches 735; Conservative 0; Mismatches 13; Indels 3; Gaps 2;

QY 2477 ATCTTGGGCTCAGGAGACTGCTTGGCCCTATTGTTGGTGGAGCTGGGGTTCGGCAA 2536
DB 1 AATCTTGGGCTCAGGAGACTGCTTGGCCCTATTGTTGGTGGAGCTGGGGTTCGGCAA 60
QY 2537 GGAATTTAGGAGCGCTTGGAGCGCAATGTTCCCGAGAGCAAGTTACAGGTGTGATGT 2596
DB 61 GGAATTTAGGAGCGCTTGGAGCGCAATGTTCCCGAGAGCAAGTTACAGGTGTGATGT 120
QY 2597 GACTTATGAGGGGAGGAGGAGCTTATGACCTTTAAGAGGCTGATGGCTCCCTTA 2656
DB 121 GACTTATGAGGGGAGGAGGAGCTTATGACCTTTAAGAGGCTGATGGCTCCCTTA 180
QY 2657 CCTGAACAGGCTGCTGCTGTGTGGAATTTCCCAAGCTGCATCAGAGACCTTCTGTG 2716
DB 181 CCTGAACAGGCTGCTGCTGTGTGGAATTTCCCAAGCTGCATCAGAGACCTTCTGTG 240
QY 2717 GCTGACACTTGGGCTTCCCATCATGTGTGCTCCGACTGAGAAAGACATTTGAACACCG 2776
DB 241 GCTGACACTTGGGCTTCCCATCATGTGTGCTCCGACTGAGAAAGACATTTGAACACCG 300
QY 2777 ATCTTGGGCTCAGGAGACTGCTTGGCCCTATTGTTGGTGGAGCTGGGGTTCGGCAA 2836
DB 301 ATCTTGGGCTCAGGAGACTGCTTGGCCCTATTGTTGGTGGAGCTGGGGTTCGGCAA 360
QY 2837 CCTCCGCGATGACGAATCATGTTGGGTTCCGGGTGAGAGAAAGCTCCGACCCGAT 2896
DB 361 CCTCCGCGATGACGAATCATGTTGGGTTCCGGGTGAGAGAAAGCTCCGACCCGAT 420
QY 2897 CGTGGGCTTCCCTGGCGTTACCAAGCATGGGACATCCCATCATGCTTGGCTCTCAA 2956
DB 421 CGTGGGCTTCCCTGGCGTTACCAAGCATGGGACATCCCATCATGCTTGGCTCTCAA 480
QY 2957 CTCCAACTACCTCGTAGGCTGTCATGTCAGAGGTCGTCCTCTTTCACAAGTA 3016
DB 481 CTCCAACTACCTCGTAGGCTGTCATGTCAGAGGTCGTCCTCTTTCACAAGTA 540
QY 3017 TTATGCTACCTGATTTCTATGTGATGCCAGAGCCATCCGGAGCATGGTGAATA 3076
DB 541 TTATGCTACCTGATTTCTATGTGATGCCAGAGCCATCCGGAGCATGGTGAATA 600
QY 3077 CATCACTGTGAGACATTTGCCATGAACCTTCTTCTCCACATCACTCGGAAGCCCC 3136
DB 601 CATCACTGTGAGACATTTGCCATGAACCTTCTTCTCCACATCACTCGGAAGCCCC 658
QY 3137 CATCAAGGTGACCTCAAGGTCGATCCGATCCCAAGGATGCCCTAGAGCCC-TGTCCTC 3195
DB 659 CATCAAGGTGACCTCAAGGTCGATCCGATCCCAAGGATGCCCTAGAGCCC-TGTCCTC 718
QY 3196 ATGATGACTCCACTTCCAGAGCGGACAA 3226
DB 719 ATGATGACTCCACTTCCAGAGCGGACAA 749

RESULT 13
LOCUS BG115714 1000 bp mRNA linear EST 30-JAN-2001
DEFINITION 602317188F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:441772.5',
ACCESSION BG115714
VERSION BG115714.1 GI:12609220
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1000)
AUTHORS NIH-MGC
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LRAM10150 row: 0 column: 21
High quality sequence stop: 704.

FEATURES
source Location/Qualifiers
1..1000

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:441772"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-Sport6;
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
oligo-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 218 a 260 c 294 g 227 t 1 others
ORIGIN

Query Match 10.9% Score 672.4; DB 10; Length 1000;
Best Local Similarity 93.5%; Pred. No. 1e-152;
Matches 747; Conservative 0; Mismatches 42; Indels 10; Gaps 4;

QY 4399 AGCCAGTGTCTTCGAGAAACCCACCGGCTGGCGGAGGTTTACAGCAAGCGGCTG 4458
DB 1 AGCCAGTGTCTTCGAGAAACCCACCGGCTGGCGGAGGTTTACAGCAAGCGGCTG 60
QY 4459 CTTGGGATTAATTCCTTGTGAATTAATTCCTTCCCGGCTCTGTCTGAGGCCATCC 4518
DB 61 CTTGGGATTAATTCCTTGTGAATTAATTCCTTCCCGGCTCTGTCTGAGGCCATCC 120
QY 4519 TGTGTTATCTGTGGTTTGTGACCCCAATGTCAGCTTGGCTGAGACTCCCGAGGT 4578
DB 121 TGTGTTATCTGTGGTTTGTGACCCCAATGTCAGCTTGGCTGAGACTCCCGAGGT 180
QY 4579 TGTGTTATCTGTGGTTTGTGACCCCAATGTCAGCTTGGCTGAGACTCCCGAGGT 4638
DB 181 TGTGTTATCTGTGGTTTGTGACCCCAATGTCAGCTTGGCTGAGACTCCCGAGGT 240
QY 4639 GAATCTAACGGGATTCACAAACCGAGTTACCACTTCCACTCCCTGATTAGATTCTGT 4698
DB 241 GAATCTAACGGGATTCACAAACCGAGTTACCACTTCCACTCCCTGATTAGATTCTGT 300
QY 4699 TCCTTGGGCTGAAACTGAATTAAGCTAATTTTGGGTACAGGTGGCAGTAGGGAACT 4758
DB 301 TCCTTGGGCTGAAACTGAATTAAGCTAATTTTGGGTACAGGTGGCAGTAGGGAACT 360
QY 4759 AGGAGGTGTGAGTGGCATTTGTCAGGAGTTTATGCCCATGAGCTGTTCTGAACCTTAC 4818
DB 361 AGGAGGTGTGAGTGGCATTTGTCAGGAGTTTATGCCCATGAGCTGTTCTGAACCTTAC 420
QY 4819 TTCTCTGAAAGTGCAGTGTGACTGTGAAGTTTCTTACCACTGAACAAAGCTAGGTTTG 4878
DB 421 TTCTCTGAAAGTGCAGTGTGACTGTGAAGTTTCTTACCACTGAACAAAGCTAGGTTTG 480
QY 4879 TCCTGTGATGACATGCTTAAGCCAGTTCCGCTTCCCTAGACCTTGCATCTGTGC 4938

Oy	481	TCCGTGATGACATAGCCTTAAAGCAATTGCCTTCCTTAGACTTTGGCATCTGHC	540
Oy	4939	TTCTATTTCTTGGAATACAGTTCTCTCTGACCTGCCCTTACCAGCTGGCTCTTTCAAG	4998
Dd	541	TTCTATTTCTTGGAATACAGTTCTCTCTGACCTGCCCTTACCAGCTGGCTCTTTCAAG	600
Oy	4999	TACSTTTTTGAAGCTGGGCTCTTTTGTGTAGCTCCACCACACCTTAGGGCTAGCTGGC	5058
Dd	601	TACTG-TTTGAAGCTGGGCTCTTTTGTGTAGCTCCACCACACCTTAGGGCTAGCTGGC	658
Oy	5059	TTTAAGGAACCTCTCCCATTTGGCAAACCGGACCGCGCGCCGACGACTGTATTCCA	5118
Dd	659	TTAAGGAACCTCTCCCATTTGGCAAACCGGACCGCGCGCGACGACTGTATTCCA	714
Oy	5119	AGGTTCCCCCCCCCAACCCAGCATCAGCTGTAGCTCCCTGCTGAGGCAAGTGTGT	5178
Dd	715	AGGTTCCCCCCCCCAACCCAGCATC---GGCGGAAGTCCGCTCTAGGCGAGGGGAT	770
Db	5179	ATGTTCCAGCAGTGGGG 5197 	
Db	771	ATGTTCCAGCAGTGGGG 789	
RESULT 14			
LOCUS	AU133571	755 bp	mRNA linear EST 24-OCT-2000
DEFINITION	AU133571 OVARC1 Homo sapiens CDNA clone OVARC1000181 5', mRNA sequence.		
ACCESSION	AU133571		
VERSION	AU133571.1	GI:10994110	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,I., Nagai,T., Sugano,S. and Isogai,T.		
TITLE	HRI human CDNA project		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel.: 81-438-52-3951 Fax: 81-438-52-3952 Email: genomics@hri.co.jp HRI human CDNA project; 5'- & 3'-end one pass sequencing; Helix Research Institute; CDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute. Location/Qualifiers 1..755 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="OVARC1000181" /clone_lib="OVARC1" /tissue_type="ovary, tumor tissue" /note="Vector:pMEBFL3"		
BASE COUNT	151 a 181 c 201 g 219 t	3 others	
ORIGIN			
Query Match 10.8%; Score 666.2; DB 9; Length 755; Best Local Similarity 97.1%; Pred. No. 3e-151; Matches 709; Conservative 0; Mismatches 16; Indels 5; Gaps 3;			
Oy	4109	TTTTAGAGATTATTTTAGAGCCAGAGCTTGCCTCGTACTTATCCCTGCTCTCCATTTC	4168
Dd	1	TTTTAGAGATTATTTTAGAGCCAGAGCTTGCCTCGTACTTATCCCTGCTCTCCATTTC	60
Oy	4169	TCTCTGTTTGAAGAAGATGAGAAGCAAGTGAAGAAAGATAGGCGCTGAAGACGC	4228

Db	61	TCTCTGTTTGAAGAGATGAGAGAGCAAGAGTGAGAAAGAAATGAGGGCTGAAGACGC	120
OY	4229	CACATCCAGATGCGCTCTTTCATCTGCTCTTCTGTTGAAACACACGTCGTGTGGGCTC	4288
Db	121	CACATCCAGATGCGCTCTTTCATCTGCTCTTCTGTTGAAACACACGTCGTGTGGGCTC	180
OY	4289	AGGCGTTTCTGAAGTCCTTCTTGGAAATGAGACAGAGATCAGAGAGGTGCACATCTGC	4348
Db	181	AGGCGTTTCTGAAGTCCTTCTTGGAAATGAGACAGAGATCAGAGAGGTGCACATCTGC	240
OY	4349	TGTGCTCTGAAGTGGTTTTCAGAGTCAGCTCTCTCCCTAGTGTAGACCAAGCCAGTGC	4408
Db	241	TGTGCTCTGAAGTGGTTTTCAGAGTCAGCTCTCTCCCTAGTGTAGACCAAGCCAGTGC	300
OY	4409	CTTGGAGAACCCACCGCGCTGGCCGGGAGATTTTACAGCAAGCGCTGCCTTGGGATTA	4468
Db	301	CTTGGAGAACCCACCGCGCTGGCCGGGAGATTTTACAGCAAGCGCTGCCTTGGGATTA	360
OY	4469	ATTCTTGTGTGAATATTCACCTTCCCGCCGCTCTGTCTGAGGCCCATCTGTATTCT	4528
Db	361	ATTCTTGTGTGAATATTCACCTTCCCGCCGCTCTGTCTGAGGCCCATCTGTATTCT	420
OY	4529	GTGGTTTTTGGACCCCTATGTACCTTGGCGTGTAGAGTCCCGGAGGTTGGTATGTC	4588
Db	421	GTGGTTTTTGGACCCCTATGTACCTTGGCGTGTAGAGTCCCGGAGGTTGGTATGTC	480
OY	4589	TAGAACAAATGGAGGCTGTGATTTGCTGTGTAACTCAGATCCAGCTTGTGAATTAAG	4648
Db	481	TAGAACAAATGGAGGCTGTGATTTGCTGTGTAACTCAGATCCAGCTTGTGAATTAAG	540
OY	4649	GCGATTTCAACACCGAGTTACCATCTTCCACTCCCTGTTAGAGATTCTGTCCTGGCT	4708
Db	541	GCGATTTCAACACCGAGTTACCATCTTCCACTCCCTGTTAGAGATTCTGTCCTGGCT	600
OY	4709	GAACTGAAATTAAGTATTTTGTGGTACAGGTCGAGTGAAGTGAAGGAGGTTG	4767
Db	601	GAACTGAAATTAAGTATTTTGTGGTACAGGTCGAGTGAAGTGAAGGAGGTTG	660
OY	4768	TGAGTG--CAATTTGTCAGGATTTTATGAGCCATGACGTG--TTTCTTGAACCTTACTTCT	4823
Db	661	TGAGTGAGGATTTTGGCAAGGATTTAGGCCATGACCTTGTTCTTGAACCTTACTTCT	720
OY	4824	GGAAGTGGAG 4833	
Db	721	TGGAAGGAG 730	
RESULT 15			
LOCUS	BF982138	951 bp	mRNA linear EST 23-JAN-2001
DEFINITION	602308959P01 NIH_MGC_88 Homo sapiens CDNA clone IMAGE:4400196 5',		
ACCESSION	BF982138		
VERSION	BF982138.1	GI:12384950	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: gga@bbs.femail.nih.gov		
	Tissue Procurement: ATCC		
	CDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		

Plate: L1AM10105 row: c column: 13
High quality sequence stop: 661.
Location/Qualifiers
1. 951
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:440196"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-Sport6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 213 a 282 c 247 g 209 t
ORIGIN

Query Match 10.7%; Score 663; DB 10; Length 951;
Best Local Similarity 93.5%; Pred. No. 2e-150;
Matches 758; Conservative 0; Mismatches 45; Indels 8; Gaps 6;

```
OY 2790 AATGAATTGAGACAGAGGCCATCCTGTCATTTGATGAGATGCTACCTCCGCCATGAC 2849
DB 1 AATGAATTGAGACAGAGGC--ATCCTGTCATTTGATGAGATGCTACCTCCGCCATGAC 58
OY 2850 GAAATCATGTTGGGTGCGGGGTGGAGAGAGTCCGGACCGCATGTGGGCTTCCCT 2909
DB 59 GAAATCATGTTGGGTGCGGGGTGGAGAGAGTCCGGACCGCATGTGGGCTTCCCT 118
OY 2910 GGGCGTTACACAGCATGGGACATCCCATAGTCTGGTCTACAACTCCAACTACTCC 2969
DB 119 GGGCGTTACACAGCATGGGACATCCCATAGTCTGGTCTACAACTACTCC 178
OY 2970 TGTAGCTGTCTCATGTGCTGACAGTGTCTCTTTCACAAGTATTATGCTTACCTG 3029
DB 179 TGTAGCTGTCTCATGTGCTGACAGTGTCTCTTTCACAAGTATTATGCTTACCTG 238
OY 3030 TATCTTATGTGATGCCCCAGGCGCATCCGGACATGGTGATGAATACATCACTGTGAG 3089
DB 239 TATCTTATGTGATGCCCCAGGCGCATCCGGACATGGTGATGAATACATCACTGTGAG 298
OY 3090 GACATTCGCATGAACCTCTTGTCTCCACATCACTCGAAGCCCCCATCAAGGTGACC 3149
DB 299 GACATTCGCATGAACCTCTTGTCTCCACATCACTCGAAGCCCCCATCAAGGTGACC 358
OY 3150 TCAAGGTGACATTCGATGCCAGATGCCCTCAGGCCCTGTCTCATGTGATGCTCCAC 3209
DB 359 TCAAGGTGACATTCGATGCCAGATGCCCTCAGGCCCTGTCTCATGTGATGCTCCAC 418
OY 3210 TTCACAGAGGGGACAGATGCATCAACTTCTGTGTAAGGTGTAAGGCTACATGCCCTC 3269
DB 419 TTCACAGAGGGGACAGATGCATCAACTTCTGTGTAAGGTGTAAGGCTACATGCCCTC 478
OY 3270 CTGTACACGAGTTCAGGGTGTGATTTGTCTTCAAGACAGCCCTGCCCCATGACAAG 3329
DB 479 CTGTACACGAGTTCAGGGTGTGATTTGTCTTCAAGACAGCCCTGCCCC--ATGAAG 536
OY 3330 ACCAAGTCTTCAAGTTCATCTAGGGGACAGCGATCTGGGGAAGAGATGAGCAGAG 3389
DB 537 ACCAAGTCTTCAAGTTCATCTAGGGGACAGCGATCTGGGGAAGAGATGAGCAGAG 596
OY 3390 GGAGGAAGATGGCTCCAGAGGTTCTAGGATGTCAGAGACCTTGGGACATCTGTGCTG 3449
DB 597 GGAGGAAGATGGCT--CCAAATTCTTAGGCA--TGCAGGACCTTGGGACATCTGTGCTG 654
OY 3450 GGTGGCCACAGAGCTCTGTGTAAGGGGACAGAGAGTGAAGGAAACCGCTGCTT 3509
DB 655 GGTGGCCACAGAGCTCTGTGTAAGGGGACAGAGAGTGAAGGAAACCGCTGCTT 714
OY 3510 TATC-TTGAAGTCCACACTGTGGCTTGAGCCCTGGGCGAGTCCCGGGGTTCCCA 3568
DB 715 TATCTTGGAGTCCACAACTGGGCTTGAGAGACCTTGGCGAAGTTCCCGGGGTTCCCC 774
OY 3569 CACAGGCACTGACTGATGATGCTTACACTGAG 3599
DB 775 AACACGG-TCTGTCTGTGATGTTTACACTGAG 804
```

Search completed: September 8, 2002, 20:08:09
Job time: 7750 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 18:50:09 ; Search time 108.24 Seconds
(without alignments)
14006.375 Million cell updates/sec

Title: US-09-809-920-3

Perfect score: 6172
Sequence: 1 GGCGGGTCCCTGAGCTGGA.....ATTAATGAGACTTAACTC 6172

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, NA: *
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/1na/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/1na/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/1na/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/1na/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	199	3.2	3003	4 US-08-915-337-1	Sequence 1, Appl
2	57.2	0.9	7218	1 US-08-232-463-14	Sequence 14, Appl
3	46	0.7	2241	2 US-08-838-2198-20	Sequence 20, Appl
4	46	0.7	2241	3 US-09-233-336A-20	Sequence 20, Appl
5	46	0.7	2241	4 US-09-233-752A-20	Sequence 20, Appl
6	46	0.7	2241	5 US-09-402-036-20	Sequence 20, Appl
7	46	0.7	2370	2 US-08-838-2198-19	Sequence 19, Appl
8	46	0.7	2370	3 US-09-233-336A-19	Sequence 19, Appl
9	46	0.7	2370	4 US-09-233-752A-19	Sequence 19, Appl
10	46	0.7	2370	5 US-09-402-036-19	Sequence 19, Appl
11	46	0.7	2403	1 US-08-471-044-30	Sequence 30, Appl
12	46	0.7	2403	2 US-08-471-044-30	Sequence 30, Appl
13	46	0.7	2403	3 US-08-463-483A-30	Sequence 30, Appl
14	46	0.7	2403	2 US-08-471-046A-30	Sequence 30, Appl
15	46	0.7	2403	2 US-08-470-566B-30	Sequence 30, Appl
16	46	0.7	2403	2 US-08-838-2198-7	Sequence 7, Appl
17	46	0.7	2403	2 US-08-469-334-30	Sequence 30, Appl
18	46	0.7	2403	3 US-09-300-529-30	Sequence 30, Appl
19	46	0.7	2403	3 US-09-233-336A-7	Sequence 7, Appl
20	46	0.7	2403	3 US-09-233-752A-7	Sequence 7, Appl
21	46	0.7	2403	4 US-09-402-036-7	Sequence 7, Appl
22	45.6	0.7	1435	4 US-09-153-804-4	Sequence 4, Appl
23	44.6	0.7	1890	6 5312912-3	Patent No. 5312912
24	43.8	0.7	16442	3 US-08-781-891-208	Sequence 208, App
25	43.6	0.7	1361	4 US-09-428-583-3	Sequence 3, Appl
26	43.6	0.7	1700	2 US-08-839-581A-1	Sequence 1, Appl
27	43.6	0.7	1700	4 US-09-023-591A-1	Sequence 1, Appl

C	28	43.6	0.7	10763	1 US-08-761-258-1	Sequence 1, Appl
C	29	43.6	0.7	10763	2 US-08-977-306-1	Sequence 1, Appl
C	30	43.4	0.7	1368	3 US-08-874-563-5	Sequence 5, Appl
C	31	43.4	0.7	1368	3 US-08-577-483-14	Sequence 14, Appl
C	32	43.4	0.7	4253	3 US-08-577-483-7	Sequence 7, Appl
C	33	43.2	0.7	4254	2 US-08-443-639-7	Sequence 7, Appl
C	34	42.4	0.7	3709	4 US-09-541-782-7	Sequence 7, Appl
C	35	42.2	0.7	3572	2 US-08-713-815A-2	Sequence 2, Appl
C	36	41.6	0.7	43280	2 US-08-804-227C-1	Sequence 1, Appl
C	37	41.2	0.7	80246	4 US-09-078-294-4	Sequence 4, Appl
C	38	40.4	0.7	51952	3 US-08-947-823-1	Sequence 1, Appl
C	39	40.2	0.7	1048	4 US-09-153-804-1	Sequence 1, Appl
C	40	40	0.6	1767	1 US-07-903-466-2	Sequence 2, Appl
C	41	40	0.6	1767	5 PCT-US93-05794-2	Sequence 2, Appl
C	42	40	0.6	3018	1 US-07-903-466-1	Sequence 1, Appl
C	43	40	0.6	3018	5 PCT-US93-05794-1	Sequence 1, Appl
C	44	40	0.6	7218	1 US-08-232-463-14	Sequence 14, Appl
C	45	39.6	0.6	939	1 US-08-285-440-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-915-337-1
Sequence 1, Application US/08915337
Patent No. 6287802
GENERAL INFORMATION:
APPLICANT: Deng, Fan & Xia
TITLE OF INVENTION: EXT2 Gene
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Smtlinkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTED Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: 536
FILING DATE: August 21, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: China - 96121928.9
FILING DATE: October 21, 1996
ATTORNEY/AGENT INFORMATION:
NAME: William T. Han
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: ATG50019
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 270-5219
TELEFAX: (610) 270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3003 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
AMTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-915-337-1
Query Match 3.2% Score 199; DB 4; Length 3003;

Best Local Similarity 57.8%; Pred. No. 3.9e-45;
Matches 396; Conservative 0; Mismatches 280; Indels 9; Gaps 2;

```
QY 2668 TCGTGGTGGTGTGATTTCTCCCAAGCTGCGATCAGAGAGACCTTCTGTGGCTTGACATTG 2727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1653 TTGTCTCTGGGAATATCAAGATAAAACCCTCGAAGATTTCTCTGTGGCCCAAAATCC 1712
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2728 GCGTCCCATCATGTGTGCTCCGACTAGAGAAACATGTTGAACCCGATTTCTACCT 2787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1713 GGGTTCATTAAAGTTTGAGAGACTGCTGMAAACAGTTAAAGTAAACCGTTCTCCCTT 1772
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2788 GGAATGAATTTGACAGAGAGCCCATCTGTCCATTGATGACAT---GCTCACTCCGCC 2844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1773 AATGATGAATTCAGACAGAGAGCTGTCTGGCCATTGATGATATCATATTATGCTACCT 1832
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2845 AATGAGAAATCATGTTGGGTTCGGGTGTGAGAGAACTGGGACCCCATCTGGGCT 2904
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1833 CTGACGAGCTGCAATTTGTTATGAGGTCTGGCGGAATTTCTCCACCGTGTGGTGGT 1892
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2905 TCCCTGGCGCTTACCAAGCATGGAGATCCCATCATGCTCCGCTCTCACTCACTCACT 2964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1893 ACCCGGATGCTGCTGCTCTCTGGACCATGATGATGAATGGAAGTATGAGCTGAGCT 1952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2965 ACTCTGTGAGCTGTCCATGTGTCTGACAGGTGCTCTTTCACAGTATTAATGCT 3024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1953 GGACGAATGAAGTTCATGATGTTGCTCAGTGGGAGCTTTTATCACAAGTATTTAAT 2012
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3025 ACCCTATTTCTATGTATGATGCCCCAGGCCATCCGGAGCATGTTGATGATATCACTCACT 3084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2013 ACCCTATTAACCTCAAAATGCTGCGGATATCAAGAACTGGGATGATGCTCATATGACT 2072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3085 GTGAGAGATTGCCATGAACTCTTGTCTCCCATCATCACTCTGGAAGCCGCCCATCAAG 3144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2073 GTGAAGATATGCTCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3145 TGACCTACAGGTGACATTCGAGATCCAGATGCC-----TCAGGCCCTGTCTCAAG 3198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2133 TAACCCCAAGAAATTTCAAGTCTCTGATGACAGCAATGATGGCTTTTACTAG 2192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3199 ATGACTCCACTTCACAGAGGGGACAGATGATCACTTCTGTGAGGTGATGAGCT 3258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2193 ACCAAACACATGATGAGAGAGGTCAGATCATCAACAGTTGGCTTCACTCTTGGGA 2252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3359 ACATGCCCTCTCTACACGAGATGATGATGCTGCTCTCAAGACAGCCCTGC 3318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2253 CCATGCCCTCTCAAGTGTGTGTAACAGCAGCTGCTGCTGTACAAAGATGACTTTC 2312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3319 CCCATGACAGACCAAGTCTTCAA 3343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2313 CTGAGAGCTGAGAGCTTCCCAA 2337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 2
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pt-F1s
US-08-232-463-14

Query Match 0.9%; Score 57.2; DB 1; Length 7218;
Best Local Similarity 1.3%; Pred. No. 2.7e-05;
Matches 5; Conservative 229; Mismatches 142; Indels 0; Gaps 0;

```
QY 4907 TTCCCTTCCCTAGACCTTGACATCTGTGCTCTATTTCTGGAATACGTTCCCT 4966
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1060 TTGCATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 4967 GACCTGCTTACACAGTGCGTCTCTCAAGTACTGTTGAAGCTGGGCTTTTGTG 5026
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1120 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5027 TAGCTCCACCCACTGTAGAGGCTAGCTGCGCTTTGAAGAACTCCCATGGCAACC 5086
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1180 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5087 GGACCCGCGCCGACAGACTGTGTTCAAGAGTCCCGCCCAACCCAGACATCA 5146
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1240 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5147 GCTGTAGCTCCCTGCTGAGGAGGTGTATGTTCCAGAGAGTGGGGCTAGAGGCC 5206
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1300 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1359
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5207 CTCTCCAGAACTTTAGTTGCGCTCTACCTGACTCTGACTGTTATCTTTAGCAG 5266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1360 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1419
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 5267 TAGCTTCTTCCCTCG 5282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1420 YYYYYYYYYYYYYYG 1435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 3
US-08-838-219B-20
Sequence 20, Application US/08838219B
Patent No. 5877012
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziele, Michael A
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M

```

? APPLICANT: Kostichka, N. Kristy
? APPLICANT: Duck, Nicholas B
? APPLICANT: Estruch, Juan J
? TITLE OF INVENTION: A No. 5877012el Class of Proteins for the
? TITLE OF INVENTION: Control of Plant Pests
? NUMBER OF SEQUENCES: 20
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: CIBA-GEIGY Corporation
? STREET: 7 Skyline Drive
? CITY: Hawthorne
? STATE: NY
? COUNTRY: USA
? ZIP: 10532
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30B
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/838,219B
? FILING DATE:
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/463,483
? FILING DATE: 06-JUN-1995
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/314,594
? FILING DATE: 09-SEP-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/218,018
? FILING DATE: 23-MAR-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/037,057
? FILING DATE: 25-MAR-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Pace, Gary M.
? REGISTRATION NUMBER: 40,403
? REFERENCE/DOCKET NUMBER: CGC 1925
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 919-541-8582
? TELEFAX: 919-541-8689
? INFORMATION FOR SEQ ID NO: 20:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2241 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "synthetic DNA encoding
? HYPOTHEICAL: NO
? US-08-838-219B-20

Query Match 0.7%; Score 46; DB 2; Length 2241;
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

OY 846 ATCCGGGAGTGGTGAGTGAAGAGCTCTCGACGTCGAGCGCCAGCGCCAGAGCTGTAC 905
DB 160 AACCAAGCAGCTGTGTAACGATCAGCGGCAAGCTGGAGCGCGTGAAGCGGACCTGTAA 219
OY 906 AGCGAGATCGCCAGCTGAATCTGAAGATCGAAGCTGTAGAGAGAGCATTTGAGAACGCC 965
DB 220 GACCTGATCGGCGCAAGCGCAACTGACACGACGAGCTGAGCAAGAGATCTTAAAGATGCC 279
OY 966 AAGCAGACCTGCTCCAGCTCAAGAAATGTATCAAGCCAGACCGAGCATTTCTACAAGAG 1025
DB 280 AACGAGCAAGCAAGCTGTGTAACGAGCTGAACAACAAGCTGAGAGCCATCAACACCATG 339
OY 1026 CT 1027
DB 340 CT 341
```

```

RESULT 4
US-09-233-336A-20
? Sequence 20, Application US/09233336A
? Patent No. 6107279
? GENERAL INFORMATION:
? APPLICANT: Warren, Gregory W
? APPLICANT: Kozziel, Michael G
? APPLICANT: Mullins, Martha A
? APPLICANT: Nye, Gordon J
? APPLICANT: Carr, Brian
? APPLICANT: Desai, Nalini M
? APPLICANT: Kostichka, N. Kristy
? APPLICANT: Duck, Nicholas B
? APPLICANT: Estruch, Juan J
? TITLE OF INVENTION: A No. 6107279el Class of Proteins for the
? TITLE OF INVENTION: Control of Plant Pests
? NUMBER OF SEQUENCES: 20
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: CIBA-GEIGY Corporation
? STREET: 7 Skyline Drive
? CITY: Hawthorne
? STATE: NY
? COUNTRY: USA
? ZIP: 10532
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30B
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/233,336A
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/838,219
? FILING DATE:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/314,594
? FILING DATE: 09-SEP-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/218,018
? FILING DATE: 23-MAR-1994
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/037,057
? FILING DATE: 25-MAR-1993
? ATTORNEY/AGENT INFORMATION:
? NAME: Pace, Gary M.
? REGISTRATION NUMBER: 40,403
? REFERENCE/DOCKET NUMBER: CGC 1925
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 919-541-8582
? TELEFAX: 919-541-8689
? INFORMATION FOR SEQ ID NO: 20:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2241 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "synthetic DNA encoding
? HYPOTHEICAL: NO
? US-09-233-336A-20

Query Match 0.7%; Score 46; DB 3; Length 2241;
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

OY 846 ATCCGGGAGTGGTGAGTGAAGAGCTCTCGACGTCGAGCGCCAGCGCCAGAGCTGTAC 905
DB 160 AACCAAGCAGCTGTGTAACGATCAGCGGCAAGCTGGAGCGCGTGAAGCGGACCTGTAA 219
OY 906 AGCGAGATCGCCAGCTGAATCTGAAGATCGAAGCTGTAGAGAGAGCATTTGAGAACGCC 965
DB 220 GACCTGATCGGCGCAAGCGCAACTGACACGACGAGCTGAGCAAGAGATCTTAAAGATGCC 279
OY 966 AAGCAGACCTGCTCCAGCTCAAGAAATGTATCAAGCCAGACCGAGCATTTCTACAAGAG 1025
DB 280 AACGAGCAAGCAAGCTGTGTAACGAGCTGAACAACAAGCTGAGAGCCATCAACACCATG 339
OY 1026 CT 1027
DB 340 CT 341
```

Db 160 AACACAGCTGCTGAACGACATCAGCGGACCTGGACGGCCTGAACGGCACCTGAAAC 219
QY 906 AGCGAATGCGCCAGCTGATCTGAAGATCGAAGCCTGTAAAGAGACATTTGAAGAGCC 965
Db 220 GACCTGATGCGCCAGCGGACCACTGAGACCGGCTGAGACAGGATCTTAAGATCGCC 279
QY 966 AAGCAGACCTGCTCCAGCTCAAGATGTATCAGCCAGACCGAGATTCCTTCAAGAG 1025
Db 280 AACGAGCAACACAGCTGCTGAACGACGTGACACAGCTGACCGCATCAACACCATG 339
QY 1026 CT 1027
Db 340 CT 341

RESULT 5
US-09-233-752A-20
Sequence 20, Application US/09233752A
Patent No. 6137033
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: A No. 6137033el Class of Proteins for the
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/233,752A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/838,219
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1925
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2241 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic DNA encoding
DESCRIPTION: VIP3A(c)"
HYPOTHETICAL: NO
US-09-233-752A-20

Query Match 0.7%; Score 46; DB 3; Length 2241;
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTGGTGATGAAGAGCTCTGACAGCTGAGGCCAAGGCCAAGAGCTGAAC 905
Db 160 AACACAGCTGCTGAACGACATCAGCGGACCTGAGACGGCGTGAACGGCAGCTGAAC 219
QY 906 AAGCAGACCTGCTCCAGCTCAAGATGTATCAGCCAGACCGAGATTCCTTCAAGAG 1025
Db 220 GACCTGATGCGCCAGCGGACCACTGAGACCGGCTGAGACAGGATCTTAAGATCGCC 279
QY 966 AAGCAGACCTGCTCCAGCTCAAGATGTATCAGCCAGACCGAGATTCCTTCAAGAG 1025
Db 280 AACGAGCAACACAGCTGCTGAACGACGTGACACAGCTGACCGCATCAACACCATG 339
QY 1026 CT 1027
Db 340 CT 341

RESULT 6
US-09-402-036-20
Sequence 20, Application US/09402036
Patent No. 6291156
GENERAL INFORMATION:
APPLICANT: Yu, Cao-Guo
APPLICANT: Estruch, Juan J.
APPLICANT: Warren, Gregory W.
APPLICANT: Desai, Nalini
APPLICANT: Kozziel, Michael
APPLICANT: Nye, Gordon
TITLE OF INVENTION: Plant Pest Control
FILE REFERENCE: S-21284C
CURRENT APPLICATION NUMBER: US/09/402,036
CURRENT FILING DATE: 2000-02-08
PRIOR APPLICATION NUMBER: PCT/EP98/01952
PRIOR FILING DATE: 1998-04-02
PRIOR APPLICATION NUMBER: 08/838,219
PRIOR FILING DATE: 1997-04-03
PRIOR APPLICATION NUMBER: 08/832,263
PRIOR FILING DATE: 1997-04-03
PRIOR APPLICATION NUMBER: 08/832,265
PRIOR FILING DATE: 1997-04-03
PRIOR APPLICATION NUMBER: 08/463,483
PRIOR FILING DATE: 1995-06-05
PRIOR APPLICATION NUMBER: 08/314,594
PRIOR FILING DATE: 1994-09-09
PRIOR APPLICATION NUMBER: 08/218,018
PRIOR FILING DATE: 1994-03-24
PRIOR APPLICATION NUMBER: 08/037,057
PRIOR FILING DATE: 1993-03-25
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 20
LENGTH: 2241
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
US-09-402-036-20

Query Match 0.7%; Score 46; DB 4; Length 2241;
Best Local Similarity 53.3%; Pred. No. 0.017;

FILED DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1925
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2370 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic DNA encoding
DESCRIPTION: VIP3A(b)"
HYPOTHETICAL: NO
US-09-233-336A-19

Query Match 0.7% Score 46; DB 3; Length 2370;
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTGGTGATGAGAGCTCTCGAGCTGAGAGCCGCAAGAGCTGAC 905
DB 160 AACGAGGAGCTGCTGACGATCATCGCGCAAGCTGAGCGGCTGAC 219
QY 906 AGCGATCGCCAGCTGATCTGAAGATCGAAGCTGTAGAGAGATTGAGACGCC 965
DB 220 GACCTGATCGCCAGGAGGACCTGACGACGAGTGAAGAGATCTTAAGATCGCC 279
QY 966 AAGCAGACCTGCTCCAGCTCAAGATGTCAATCAACCCAGACGATTCCTACAGAG 1025
DB 280 AACGAGGAGAGAGCTGCTGACGATGACGAGCAACAGCTGAGCGCATCAACCATG 339
QY 1026 CT 1027
DB 340 CT 341

RESULT 9
US-09-233-752A-19
Sequence 19, Application US/09233752A
Patent No. 6137033
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: A No. 6137033el Class of Proteins for the
TITLE OF INVENTION: Control of Plant Pests
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/233,752A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/838,219
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1925
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2370 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "synthetic DNA encoding
DESCRIPTION: VIP3A(b)"
HYPOTHETICAL: NO
US-09-233-752A-19

Query Match 0.7% Score 46; DB 3; Length 2370;
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTGGTGATGAGAGCTCTCGAGCTGAGAGCCGCAAGAGCTGAC 905
DB 160 AACGAGGAGCTGCTGACGATCATCGCGCAAGCTGAGCGGCTGAC 219
QY 906 AGCGATCGCCAGCTGATCTGAAGATCGAAGCTGTAGAGAGATTGAGACGCC 965
DB 220 GACCTGATCGCCAGGAGGACCTGACGACGAGTGAAGAGATCTTAAGATCGCC 279
QY 966 AAGCAGACCTGCTCCAGCTCAAGATGTCAATCAACCCAGACGATTCCTACAGAG 1025
DB 280 AACGAGGAGAGAGCTGCTGACGATGACGAGCAACAGCTGAGCGCATCAACCATG 339
QY 1026 CT 1027
DB 340 CT 341

RESULT 10
US-09-402-036-19
Sequence 19, Application US/09402036
Patent No. 6291156
GENERAL INFORMATION:
APPLICANT: Estruch, Juan J.
APPLICANT: Yu, Cao-Quo
APPLICANT: Warren, Gregory W.
APPLICANT: Desai, Nalini
APPLICANT: Kozziel, Michael
APPLICANT: Nye, Gordon
TITLE OF INVENTION: Plant Pest Control
FILE REFERENCE: S-21284C
CURRENT APPLICATION NUMBER: US/09/402,036
CURRENT FILING DATE: 2000-02-08

```
; PRIOR APPLICATION NUMBER: PCT/EP98/01952
; PRIOR FILING DATE: 1998-04-02
; PRIOR APPLICATION NUMBER: 08/838,219
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/832,263
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/832,265
; PRIOR FILING DATE: 1997-04-03
; PRIOR APPLICATION NUMBER: 08/463,483
; PRIOR FILING DATE: 1995-06-05
; PRIOR APPLICATION NUMBER: 08/314,594
; PRIOR FILING DATE: 1994-09-09
; PRIOR APPLICATION NUMBER: 08/218,018
; PRIOR FILING DATE: 1994-03-24
; PRIOR APPLICATION NUMBER: 08/037,057
; PRIOR FILING DATE: 1993-03-25
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 2370
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
; US-09-402-036-19

Query Match      0.7%; Score 46; DB 4; Length 2370:
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTCGCTGAGTGAAGAGCTCTGACGTGGAGGCCAAGCGCAAGAGCTGAAC 905
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 160 aaccagagagctgtcgaacgcacacgcagcgcaagctgagcaagcgcaagcctggaac 219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 906 AGCGAGATCGCCCAAGCTGAATCTGAAGATCGAAGCTGTGAAGAGCAATTGGAAGCGCC 965
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 220 gacctgtcgcgcagcagcgaacacgacacgcagctgagcaagagatccttaagatcgcc 279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 966 AAGCAGGACCTGCTCCAGCTCAAGAATGTGATGAGCCAGACCGACGCTTCTCTCAAGGAG 1025
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 280 aacgagagagaacagagtgctgcaacgacgtagaacaagaagctgagcgcacatacaacatg 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1026 CT 1027
    ||
    340 ct 341

RESULT 11
US-08-471-033-30
; Sequence 30, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desal, Nalin M
; APPLICANT: Koslichka, N. Kistly
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLv3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2403 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 11..2389
; OTHER INFORMATION: /note= "maize optimized DNA"
; OTHER INFORMATION: sequence encoding VIP3a(a)"
; US-08-471-033-30

Query Match      0.7%; Score 46; DB 1; Length 2403:
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTCGCTGAGTGAAGAGCTCTGACGTGGAGGCCAAGCGCAAGAGCTGAAC 905
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 176 AACCAGAGCTGCTGTAACGATCAGCGGCAAGCTGGACGCGTGAAGAGCGACGCTGAAC 235
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 906 AGCGAGATCGCCCAAGCTGAATCTGAAGATCGAAGCTGTGAAGAGCAATTGGAAGCGCC 965
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 236 GACCTGATCGCCAGGCAACCTGAACACCGAGCTGAGCAAGAGATCCTTAAGATCGCC 295
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 966 AAGCAGGACCTGCTCCAGCTCAAGAATGTGATGAGCCAGACCGACGCTTCTCTCAAGGAG 1025
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 296 AAGCAGGAGAACGAGTGTGTAAGAGAGTGAACAACAAGCTGAGCGCATCAACACCATG 355
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1026 CT 1027
    ||
    356 CT 357

RESULT 12
US-08-471-044-30
; Sequence 30, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desal, Nalin M
```


APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,044
FILING DATE: 06-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: 40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8689
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 2403 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHEICAL: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11..2389
OTHER INFORMATION: /note="maize optimized DNA"
OTHER INFORMATION: sequence encoding VIP3A(a)"
US-08-471-044-30

Query Match 0.7%; Score 46; DB 2; Length 2403;
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTGGTGAAGAGCTCTGACCTGAGGCGCAAGCGCAAGAGCTGAAC 905
DB 176 AACACAGCAGCTGCTGAACATCAGCGGCAACCTGAGCGGCTGAACGCGCTGAAC 235
QY 906 AGCGAATGCGCAAGCTGAATCTGAAGATCGAAGCTGTAGAAGAGCATTTGAGAAAGCC 965
DB 236 GACCTATGCGCCAGGAGCAACCTGAGACACCGAGCTGAGCAAGAGAGATCTTAAGATCGCC 295
QY 966 AAGCAGAGCTGCTCCAGCTCAAGATGTATCAGCCAGCAGCAGCATTCCTACAGGAG 1025
DB 296 AACGAGCAACACAGCTGCTGAACAGCAGTGAACACAAACAGCTGAGCCATCAACACCAAG 355

QY 1026 CT 1027
DB 356 CT 357

RESULT 13
US-08-463-483A-30
Sequence 30, Application US/08463483A
Patent No. 5849870
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Kozief, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,483A
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 2403 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHEICAL: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11..2389
OTHER INFORMATION: /note="maize optimized DNA"
OTHER INFORMATION: sequence encoding VIP3A(a)"
US-08-463-483A-30

Query Match 0.7%; Score 46; DB 2; Length 2403;
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 846 ATCCGGAGTGTGAGTGAAGAGCTCTCGAGCTGGAGGCCAAGCCCAAGAGCTGAAC 905
176 AACGAGAGCTGTGTAACATCAGCGCGCAAGCTGGAGCGGCGTGAACGGCAGCTGAAC 235
QY 906 AGCGAGATCGCCAGCTGAATCTGAAGATCGAAGCTGTGAAGAAGAGCATTTGAGAAGCC 965
236 GACCTGATCGCGCCAGCGGCAACCTGAACACCGAGCTGAGCAAGAGATCTTAAGATCGCC 295
QY 966 AAGCAGAGCTGTCCAGCTCAAGATGTCTATCAGCCAGCAGAGCATTCCTACAGAGG 1025
296 AACGAGCAGAACCAAGGTGTGTAACAGCAGCTGAACACAAAGCTGAGCGCATCAACACCATG 355
Db 1026 CT 1027
356 CT 357
RESULT 14
US-08-471-046A-30
Sequence 30, Application US/08471046A
Patent No. 5866326
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5866326artls Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471.046A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLv4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 2403 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHEICAL: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11..2389
OTHER INFORMATION: /note= "maize optimized DNA
US-08-471-046A-30
Query Match 0.7%; Score 46; DB 2; Length 2403;
Best Local Similarity 53.3%; Pred. No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 846 ATCCGGAGTGTGAGTGAAGAGCTCTCGAGCTGGAGGCCAAGCCCAAGAGCTGAAC 905
176 AACGAGAGCTGTGTAACATCAGCGCGCAAGCTGGAGCGGCGTGAACGGCAGCTGAAC 235
Db 1026 CT 1027
356 CT 357
QY 906 AGCGAGATCGCCAGCTGAATCTGAAGATCGAAGCTGTGAAGAAGAGCATTTGAGAAGCC 965
236 GACCTGATCGCGCCAGCGGCAACCTGAACACCGAGCTGAGCAAGAGATCTTAAGATCGCC 295
QY 966 AAGCAGAGCTGTCCAGCTCAAGATGTCTATCAGCCAGCAGAGCATTCCTACAGAGG 1025
296 AACGAGCAGAACCAAGGTGTGTAACAGCAGCTGAACACAAAGCTGAGCGCATCAACACCATG 355
Db 1026 CT 1027
356 CT 357
RESULT 15
US-08-470-566B-30
Sequence 30, Application US/08470566B
Patent No. 5872212
GENERAL INFORMATION:
APPLICANT: Warren, Gregory W
APPLICANT: Koziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalin M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
TITLE OF INVENTION: No. 5872212el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5872212artls Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470.566B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV4 - SOLV4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 2403 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "Synthetic DNA"
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 11..2389
OTHER INFORMATION: /note= "maize optimized DNA"
OTHER INFORMATION: sequence encoding VIP3A(a)"
US-08-470-566B-30

Query Match 0.7%; Score 46; DB 2; Length 2403;
Best Local Similarity 53.3%; Pred No. 0.017;
Matches 97; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 846 ATCCGGAGTGGTGAGTGAAGAGCTCTGCAGCTGAGGCCAAGCGCAAGAGCTGAAC 905
Db 176 AACGAGCAGCTGCTGAACGATCAGCGCAAGCTGAGCGGCTGGAACGCGCTGAAC 235
QY 906 AGCGAGATCCGCCAAGCTGAATCTGAAGTCGAAGCCTGTAAAGAAAGCATTGAGAACGCC 965
Db 236 GACCTGATCGGCCAAGGCAACCTGAACGAGCTGAGCAAGAGATCTTAAGATCGCC 295
QY 966 AAGCAGGACCTGCTCAGCTCAGATGTCATCAGCCGAGCCAGCATTCCTACAGAGAG 1025
Db 296 AACGAGCAGAACGAGCTGATGACGACGTGACACAAAGCTGAGCGCCATCAACACCATG 355
QY 1026 CT 1027
Db 356 CT 357

Search completed: September 8, 2002, 22:22:30
Job time: 12741 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 8, 2002, 18:03:19 ; Search time 590.89 Seconds
(without alignments)
17933.619 Million cell updates/sec

Title: US-09-809-920-3
Perfect score: 6172
Sequence: 1 GCGGGCTCCTGAGCTGAA.....ATMAATGAGACTTTACTC 6172

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0
Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*

1:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDSL/gcgdata/hold-geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	6172	100.0	6172	21 AAA12734
2	5833	94.5	5854	20 AAV72378
3	2459.6	39.9	3479	21 AAA12733
4	2347.4	38.0	3198	22 AAF26844
5	2212.4	35.8	2253	21 AAC59000
6	882.2	14.3	1599	22 AAF26843
7	596.4	9.7	598	22 AAF68305
8	545.6	8.8	3595	23 ABL26683
9	545.6	8.8	6105	23 ABL26682

10	310.4	5.0	316	14 AA061373	Human brain Express
11	267.8	4.3	756	20 AAV8968	EST clone IE352.
12	207.2	3.4	353	20 AAV90056	EST clone C21. Ho
13	200.6	3.3	3008	21 AAF21849	Human breast and o
14	200.6	3.3	3175	20 AA088224	Human Ext-2 CDNA.
15	199	3.2	3003	19 AAV19375	Hereditary multipl
16	169.4	2.7	3183	20 AA088223	Human Ext-1 CDNA.
17	166.2	2.7	2698	23 ABL09865	Drosophila melanog
18	166.2	2.7	3642	20 AA088225	Drosophila melanog
19	163	2.6	3817	23 ABL16129	Drosophila melanog
20	160.2	2.6	2965	24 ABL99788	Mouse ischemic co
21	159	2.6	176	16 AAT24258	Human gene signatu
22	137.4	2.2	4733	23 ABL10422	Drosophila melanog
23	137.4	2.2	4820	23 ABL09864	Drosophila melanog
24	90.2	1.5	60747	23 ABL16128	Drosophila melanog
25	68.4	1.1	617	23 AAS94481	DNA encoding novel
26	52.4	0.8	6976	22 AAK52949	Human polynucleoti
27	52.4	0.8	7569	22 AAK51965	Human polynucleoti
28	51	0.8	51	22 AAT17426	Human silent SNP c
29	51	0.8	51	22 AAT17428	Human silent SNP c
30	50.2	0.8	23449	21 AAT35393	Maize amylose-exte
31	50	0.8	50	22 AAT17430	Human silent SNP c
32	49.8	0.8	905	21 AAT77427	Human cancer assoc
33	49.4	0.8	51	22 AAT17429	Human silent SNP c
34	49.4	0.8	51	22 AAT17429	Human silent SNP c
35	49.2	0.8	42299	22 AAK68932	Human immune syste
36	48	0.8	8323	24 ABL32058	Human immune syste
37	47.8	0.8	5987	24 ABL33563	Human immune syste
38	47.2	0.8	10732	21 AAA10594	Gene encoding a su
39	47	0.8	8136	24 ABL32554	Human immune syste
40	47	0.8	99960	21 AAK50905	Human TBC-1 patita
41	46.8	0.8	7516	24 ABL32060	Human immune syste
42	46.8	0.8	17967	24 ABL33014	Human immune syste
43	46.6	0.8	18301	22 AAT198944	Human excretory re
44	46.6	0.8	18301	22 AAT198944	Human kidney relat
45	46.4	0.8	985	22 AAK51884	Human polynucleoti

ALIGNMENTS

RESULT 1
ID AAA12734 standard; cDNA: 6172 BP.

AC AAA12734
XX 25-JUL-2000 (first entry)

DE cDNA encoding human TREX protein.

KM Tumour necrosis factor receptor-associated Factor; TRAF;
TRF-protein-interacting hereditary multiple exosces protein; TREX;

KW signal mediator; tumour necrosis factor receptor;
CD40 mediated signal transduction; TRAF protein; cancer;
hereditary multiple exosces; autoimmune disease; ss.

OS Homo sapiens.

XX Key Location/Qualifiers
FT CDS 594..3353
FT /tag-a
FT /product- "tumour necrosis factor receptor-associated
Factor (TRAF) protein-interacting hereditary
multiple exosces (TREX) protein"

XX W0200018959-A1.

XX 06-APR-2000.

XX 17-SEP-1999; 99WO-US21654.

XX 17-SEP-1998; 98US-0156191.

Applied

XX (UYCO) UNIV COLUMBIA NEW YORK.
 XX Sato T;
 XX WPI: 2000-293180/25.
 DR P-PSDB: AAY84669.
 XX New nucleic acid encoding Tumour necrosis factor Receptor-Associated
 PT Factor (TRAF) protein-interacting hereditary multiple exstoses (TRES)
 PS protein, useful in the diagnosing cancer -
 PS Claim 21; Fig 8A-B; 161pp; English.
 XX
 CC The present sequence encodes a human tumour necrosis factor receptor-
 CC associated factor (TRAF) protein-interacting hereditary multiple
 CC exstoses (TRES) protein. TRES is a signal modulator which bridges
 CC between the tumour necrosis factor (TNF) receptor and CD40 mediated
 CC signal transduction. Anti-TRES antibodies are useful for treating an
 CC abnormality in a patient by inhibiting binding of a TRES protein and
 CC a TRAF protein (that is, TRAF 2, TRAF 3 or TRAF 5). The abnormality
 CC is cancer, a hereditary multiple exstosis or an autoimmune disease. The
 CC cancer is colon cancer, gastric cancer, human head and neck squamous
 CC cell carcinoma, prostate carcinoma, breast cancer, thyroid cancer,
 CC oesophageal cancer, lung cancer, colorectal cancer, ovarian cancer,
 CC papillary bladder cancer, osteosarcoma, chondrosarcoma, liposarcoma,
 CC giant cell tumour, Ewing sarcoma, or other malignant tumours.
 XX
 S0 Sequence 6172 BP; 1280 A; 1681 C; 1692 G; 1519 T; 0 other;

Query Match 100.0%; Score 6172; DB 21; Length 6172;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 6172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCGGGTCCCTGAGCTGGAAGCCGAGAGCAAGCCCTGAGCTTCACTCTTCAAGAGT 60
 Db 1 ggcgggtccctgaagctgaagccgagagcaagccctgaggttcaacttcaagaagt 60
 QY 61 GGTGGCTGAGGTGTAATGCTACACAACTCAGAGAGAGAGGCTCTGAACACATGCG 120
 Db 61 cgtggctgaagtgtaatgctacacaaactcagagagagagggctcctgaacacatgcg 120
 QY 121 CTGATTTGGCAAGGCATATAAGAGCTGGCATTTATTTCTGCTTAACCTTTACT 180
 Db 121 ctgatttggcgaagcgcataataagagctggcatTTTATTTCTGCTTAACCTTTACT 180
 QY 181 GTATTAAGCTGTAATGACACTATGCAATTTGTTGGTCAGCAAAACCAAGAACAGAGC 240
 Db 181 gtataagctgtatgacactatgcatattgttgcagcaaaaaccaagaagaagcgc 240
 QY 241 TATGCAATTTGAAGAGCTGCTGATTTCCAGGGTGTTCCTGGGTTTATCATCAGG 300
 Db 241 tatgcaatttgaagagctgctgattccaggggtgttccctggggttattcatcagg 300
 QY 301 TACCTCTCCCTTTCATCTCAGCAAGATGCGACCTTTATGCTTATGATTAAGATTAA 360
 Db 301 tactctctcccttctcatctcagcaagaatgycacactttatcgtttgataaagaattaa 360
 QY 361 GGACATTTCTTTGGTCAACAGCCAGAACTTAAATGTGTGAATAGGGTCAGAGACA 420
 Db 361 ggaacatttcttggcaaacagccagaaacttaaatctgtggaatagggtcagagaca 420
 QY 421 TTTCAGCTGAGCTGAGGAAATGAATGTTCAATTTATTTGGTCTTGTCTGGGAGC 480
 Db 421 tttcagctgagctgaggaatgaatgttcaattttatTTTGGTCTTGTCTGGGAGC 480
 QY 481 ACACTAACTCTTGTGAAGAGTGTCACTGAACAGAGATGTTTGTGGAATACACACC 540
 Db 481 acaactaactcttgtgaagagtgtcactgaacagagatgTTTGTGGAATACACACC 540
 QY 541 ATGTTATGCGGAGTGACCCGACGCTGATTTGGGGGCGAGGCTGCAGAGACTATGACAG 600
 Db 541 atgttatgCGGAGTGACCCGACGCTGATTTGGGGGCGAGGCTGCAGAGACTATGACAG 600

Db 541 atgttatgCGGAGTGACCCGACGCTGATTTGGGGGCGAGGCTGCAGAGACTATGACAG 600
 QY 601 GCTATACATGCTGCGGAATGGGGCGGAGGAGAGAGTACAGCTGATCTGCGCT 660
 Db 601 gctatacatgctgCGGAATGGGGCGGAGGAGAGAGTACAGCTGATCTGCGCT 660
 QY 661 GGTCCAAACCGCATCCGCCCTCAGCTGCTCAGCTTCACGCTTTTGTATCTGGTCTCT 720
 Db 661 ggtccaaacCGCATCCGCCCTCAGCTGCTCAGCTTCACGCTTTTGTATCTGGTCTCT 720
 QY 721 TCCCGCTCATGCGCCACTTTTACCTACACACTGATGATAGGCTGATGAGCGCAAGC 780
 Db 721 tcccgctcatgCGCCACTTTTACCTACACACTGATGATAGGCTGATGAGCGCAAGC 780
 QY 781 GGATTTTGTGTCGCCGGGTGGGGAAAGAGCTGTGCAAGTGAAGACAGTGTGATCTGT 840
 Db 781 ggattttgtgtgTCGCCGGGTGGGGAAAGAGCTGTGCAAGTGAAGACAGTGTGATCTGT 840
 QY 841 GCGCAGTCGGGAGTGTGAGTGAAGAGAGCTCTGCAAGCTGAGAGCCCAAGCCCAAGAGC 900
 Db 841 gcgcagtcgggagTGTGAGTGAAGAGAGCTCTGCAAGCTGAGAGCCCAAGCCCAAGAGC 900
 QY 901 TGAACAGCGAGATGCGCAAGCTGAATCTGAATGCAATGCAACCTGTAAAGAGACATTGAGA 960
 Db 901 tgaaacagcgagATGCGCAAGCTGAATCTGAATGCAATGCAACCTGTAAAGAGACATTGAGA 960
 QY 961 ACGCCAAAGCAGAGACCTGCTCCAGCTCAAGAAATGTATCAGCCAGACCGCATTCCTACA 1020
 Db 961 acgccaagcagAGACCTGCTCCAGCTCAAGAAATGTATCAGCCAGACCGCATTCCTACA 1020
 QY 1021 AGGAGCTCATGCGCCCAAGAACAGCCCAAGCTGTCCCTGCCCCATCCGACTGTCCAGAGA 1080
 Db 1021 aggagctcatgCGCCCAAGAACAGCCCAAGCTGTCCCTGCCCCATCCGACTGTCCAGAGA 1080
 QY 1081 AGGAGATGCGCGGCTCCCTCCCGAAGGCGCACTGCGGGGCTGCGGCTACACACTGCT 1140
 Db 1081 aggagatgCGCGGCTCCCTCCCGAAGGCGCACTGCGGGGCTGCGGCTACACACTGCT 1140
 QY 1141 TTGATTAATTCGTTGCCCTCTCACTCTGCTTCCCGTCTACGTCATGACAGTACAGC 1200
 Db 1141 ttgattaatTCGTTGCCCTCTCACTCTGCTTCCCGTCTACGTCATGACAGTACAGC 1200
 QY 1201 AGTTGTCTTTGGCAGCTACCTGATCCCTTGTGTCAACAGAGCTTTTACAGCGACAGAC 1260
 Db 1201 agttgtctTTTGGCAGCTACCTGATCCCTTGTGTCAACAGAGCTTTTACAGCGACAGAC 1260
 QY 1261 GAGCTAAGCTTTATGTTTACAGAAATGACAGACTGCGCTTATGATGATATGAGG 1320
 Db 1261 gagctaaGCTTTATGTTTACAGAAATGACAGACTGCGCTTATGATGATATGAGG 1320
 QY 1321 GAGAGATGACAGAGCCCGTGTGCTGCGGCTGCTGAGACTTGAGAGACAGTGTATTCCC 1380
 Db 1321 gagagatgacAGAGCCCGTGTGCTGCGGCTGCTGAGACTTGAGAGACAGTGTATTCCC 1380
 QY 1381 TGCCACAATGGCGGAGCGATGAGACAAACATGTCATCATATATCTGTACAGTAAGTCAG 1440
 Db 1381 tgccaacaatGGCGGAGCGATGAGACAAACATGTCATCATATATCTGTACAGTAAGTCAG 1440
 QY 1441 ATACAGAAACCTTCTATATACCTGATACCTGACCGCTGACATGAGGCGCAGACACT 1500
 Db 1441 atacaagaAACCTTCTATATACCTGATACCTGACCGCTGACATGAGGCGCAGACACT 1500
 QY 1501 TCTAAGCTGTCCAGTACAGACCTGGCTTTGACTTGTGATACCCGCTGGTCCATGCCA 1560
 Db 1501 tctaagctgtCCAGTACAGACCTGGCTTTGACTTGTGATACCCGCTGGTCCATGCCA 1560
 QY 1561 TGTCTGAGCCCACTTATATGGAATCCACACAGGTGCGGTGAAGCGCAATATCTCT 1620
 Db 1561 tgtctgagCCCACTTATATGGAATCCACACAGGTGCGGTGAAGCGCAATATCTCT 1620
 QY 1621 TCACCTTCCAGGGGAGAAATGAGTCTGTGAGAGTCTTACCTTACAGGAGGCGCGCTCT 1680
 Db 1621 tcaccttccAGGGGAGAAATGAGTCTGTGAGAGTCTTACCTTACAGGAGGCGCGCTCT 1680

QY 1681 TCGAAGAGAATGAGGGGACCTTCGCCGACTAGATGACCGGATCATTTGCCACC 1740
|||||
Db 1681 tcgaagagagaaatcgagagggagacccctccgcgcgactacgatlgaacggatcatlgtccaccc 1740
QY 1741 TGAAGCGGTGACAGACAGCAACCTGGATCAGTCTGTGTGAATTTCACGCAAAAACC 1800
|||||
Db 1741 tgaagcggtgacagacagcaaacctgtagtaccgtgtccgtgtgaatttcaacctgcaaaaacc 1800
QY 1801 AGGCCAAACCCGACTGGCCGACTGAGTGGGACTGTGTGAGAGCGGGAGACCGGCTTGG 1860
|||||
Db 1801 aggccaaacccgactggccgactgagtgaggactgtgtgagagcgggagacccgcttgg 1860
QY 1861 AATTGCTGAAGCTCTCCACCTTCGCCCTCATCTTACCCCCGGGGACCCCTGCTGGTTA 1920
|||||
Db 1861 aattgctgaagctctccaccttcgccctcatctttacccccggggacccctgctggttta 1920
QY 1921 TTTTCCTTGGGTGTGCACACAGCGCTCTCGAACCCCTGGAAAGTCGGTCCCGCTGG 1980
|||||
Db 1921 ttttccttgggtgtgcacacagcgctctcgaaagccctggaagtcggtcccgctgg 1980
QY 1981 TGGTGGGGAGCAGGTCCACGCTTCCCTACAGGACATGCTGAGTGGAAAGAGCGGCC 2040
|||||
Db 1981 tggctggggagcaggtccacgcttccctacaggacatgctcgcaagtgaagagcggtccc 2040
QY 2041 TGGTGTGCGCAAAAGCCTGTGTTACGAGGTTGATTTCTGCTCAGAAAGCCTCTCGATA 2100
|||||
Db 2041 tgggtgtgcccataagcctgtgttacccgaggttcaattctcgtcagaaagcctctcgata 2100
QY 2101 GTACACCTCCGTGGCTATGAGCGCGGCAAGCGCGCTTCTCTGGGAGACTTACTTCTCCACTG 2160
|||||
Db 2101 gtacacctccgtggctatgagcggcgcaagcgcttctctgggagacttacttctccactg 2160
QY 2161 CTACAGATATTTTAAATACGCTGCTGCTATGATTAAGACTCCCATCCAGTCCAGCCG 2220
|||||
Db 2161 ctacagatatttttaataacccgtgtgtgtatgtattgagactcgatccagatcccgagccg 2220
QY 2221 CTCGCCATCCGGGAGAGAGCGCGCAGCTGAGATCCCCACCGTTACAGGAGCGCGTGGAA 2280
|||||
Db 2221 ctccgatccgggaagagcgcgagctgaga tccccacggttcaggaagagcgctgga 2280
QY 2281 CTGACCCCAACATGCTGACAAAGGGGACCTGACACTGGGGCCAGTGGAAACGAGCCG 2340
|||||
Db 2281 ctgaccccaacatgctgacaaaggagccctggaacctgggagcagtgagaacgagacgcgc 2340
QY 2341 CCGACGCTACCCAGATACCTCCGCAATTTCACTGACTGTACATGATGATTTTACCGCA 2400
|||||
Db 2341 ccgacgctacccagatacctccgcgaatttcaactctgactgtcaactgttllacccgca 2400
QY 2401 GCTGGAACGTGTGCTCAAGGGCCTTCCATCTTTCCCCACACTCCCTTTGACCGTGTG 2460
|||||
Db 2401 gctggaaactgtgtctcaaggccttccacttcttccccacactcccttggaccctgtgt 2460
QY 2461 TGGCCCTAGAGGCCCAATTTCTGGGCTCAGGAGACTGGCTTTGGGCTATTTGGTGTGAG 2520
|||||
Db 2461 tggccctagagggccaatttcttgggctcaggagactggcttccgctcatgtgtgtgag 2520
QY 2521 CTGGGGGTTTGGCAAGAAATTTACAGACGCGTGGAGCAATGTATCCCGAGAGAGT 2580
|||||
Db 2521 ctgggggtttggcaagaaatttacagacgcggttgagcaaatgtatcccgagagagct 2580
QY 2581 TCACGGGTGATGATTGACTTATGAGCGGAGAGAAAGTCTTATGAACCTTTAGAGAGGC 2640
|||||
Db 2581 tcacgggtgatgattgacttattagcgggagagaaagtcttattgaactctttagagggc 2640
QY 2641 TGAATGGCTTCCTTACCTAACAAGTCTGTGTGTGTGAATTTCTCCAAAGCTGCCAT 2700
|||||
Db 2641 tgaatggcttccttaccttaacaagtctgtgtgtgtgaattttctccaaagcttgccat 2700
QY 2701 CAGAGGACCTTCTGTGCTGACATTTGGCGTTCCATCATGTGTGTCTGACTAGAGAGA 2760
|||||
Db 2701 cagaggaacctctgtgcttgacatgtgcttcccatcagtggtgtgtcgttactgtagaaga 2760

QY 2761 ACAGTTGAACAACGATTCTTACCTCGAATGAAATTTGAGACAGAGCCATCTCTCCA 2820
|||||
Db 2761 acagttgaaacaacgatttcttaccctgtgaatltgaaagagggccatctctgtcca 2820
QY 2821 TTGATGACGATGCTCAACCTCCGCCATGACGAATCATGTTTGGGTTTCCGGTGTGGAGAG 2880
|||||
Db 2821 ttgatgacgatgctcaacctccgccatgacgaatcatgtttgggtttccgggttgagag 2880
QY 2881 AAGCTGGAGCCGATCGTGGGCTCCCTGGCGTTTACACAGCATGGAGATCCCATC 2940
|||||
Db 2881 aagctggagccgatcggtggcttccctggcgtttaccagcatgagagatccccatc 2940
QY 2941 AGTCGTGCTCTACAACTCCAACTACTCTGTGAGCTGTTCATGTGTGAGAGTGTG 3000
|||||
Db 2941 agtcgtgctctacaactccaactcactctgtgagctgttccatcgtgtcgaagtgctg 3000
QY 3001 CCTTCTTTCAAGATATTATGCTTACCTGTATTTCTTATGTGATGCCCCAAGCCATCCGG 3060
|||||
Db 3001 ccttcttcaagatattatgcttacctgtatgtttcttattgtgagtcgcccaagccatccgg 3060
QY 3061 ACATGTGATGATATCATCACTGATGAGACATTTGCAATGAACTTCTTGTCTCCACA 3120
|||||
Db 3061 acatgtgatgatatacatcactgatgagacatttgcaatgaacttcttgtctccaca 3120
QY 3121 TCACTGGAAAGCCCCCATATCAAGGTACCTCAGGTGAGACATTCGATGCCAGATGCC 3180
|||||
Db 3121 tcaactggaaagcccccatatcaaggttgaaccttcaagtgtagaatctcgaatgccagatgcc 3180
QY 3181 CTCAGGCGCTGTCTCATGATGACTCCACTTCACAGCGGGCAAGTGTATCAACTCT 3240
|||||
Db 3181 ctcagcgctgtctcatgatgactccacttcacagcgggcaagtgtatcaactct 3240
QY 3241 TCGTGAAGTGTACGGCTACATGCCCTCCTGTACACGAGTTTCAGGGTGTGATTTCTGC 3300
|||||
Db 3241 tcgtgaagtgtacggctacatgccctcctgtacacgagtttcaggggtgtgatttctgc 3300
QY 3301 TCTTCAAAGACAGCGCTGCCCATGACAAACAAAGTGTCTTCAAGTTCTATGTAGGGCAGC 3360
|||||
Db 3301 tcttcaaagacagcgctgcccatgacaaacaaagtgtcttcaaagttcatctaaagggtgagc 3360
QY 3361 GCAAGGCTGTGGGAAAGAGATGAGCAGAGGAGAGATGTGCCCAAGGTTCTTAGGCA 3420
|||||
Db 3361 gcaaggctgtgggaaagagatgagcagagagagaaatgagttcccaagttcttaggca 3420
QY 3421 TTGCAGGACCTTGGGCAACATCTGCTGGTGGGCCAGACCTTCTGTGGAAGGGCAG 3480
|||||
Db 3421 ttgcaggaccttgggcaacatctgtgtgtgtggccagacaccttctgtgaagggtcag 3480
QY 3481 CAGAGAGTGAAGAAACCGCTCCCTTATCTTGAATGACGCACACTGGGCTGGAG 3540
|||||
Db 3481 cagagagtgaagaaacgctcccttattcttgaatgacgcacactgggctggag 3540
QY 3541 CCCGTGGCGGAGTCCCGGGGTTCCACAGAGGCACTGACGATGATGCTTACAGTAGG 3600
|||||
Db 3541 cccgtggcgagttcccggttcccaacagggacgtgagatgacttcaacgtgaag 3600
QY 3601 ACTGTGGCAGCTGTGAGAGTCACTACACGTTTCTGTACGCCAGAGACAGTGTGCTG 3660
|||||
Db 3601 actgtggcagctgtgagagtcaactcaacacctgtgtgaagcccaagacagctgtgtgtg 3660
QY 3661 GTTTTTCATTAATAACAATATATGATGATTTTAAAGAGAAATTTTCAAGATTTGGC 3720
|||||
Db 3661 gtttttcatatlaaacaactatlaagatlaattlaaagagaaagtttcaagatltggc 3720
QY 3721 ATTCAAGGCTTATTTATATATGATGCTGTATATAATCATGSCACACTGACATACA 3780
|||||
Db 3721 attcaaggcttatttataataatgctgtatataatcatgscacactgacataca 3780
QY 3781 TATATATTTTGGCTGGGAGAGTGTAGTGTGCTTCTTCTTAAAGGAGAGCCGCGAGCG 3840
|||||
Db 3781 tataatatlttggctgggagagtgtagtlttgcttccaaagagagcgcgagggc 3840
QY 3841 TCCTTTGTCTGTATTTCTGGCGGAGATGGTCTTGCCCTTGTGTCACTGGCTTATCCTTA 3900
|||||

QY 6061 GAACCTTTGATATGATGCGACGTCATCCACCCACCCCTTTTGGTCACGACACATG 6120
|||||
Db 6061 gaaacttgcataatcaccgacgacacccaccccttctgacagacacatg 6120
QY 6121 GTCTTACATGATTTTGTAAAAAATAAATAATGAGACCTTTAACTC 6172
|||||
Db 6121 gtcttaccatcgatcttctgtaaaaaataaataatgagacttaacc 6172

RESULT 2
AAV72378
ID AAV72378 standard; cDNA; 5854 BP.

AC AAV72378
XX 29-JUL-1999 (first entry)
XX Human EXOSTOSIN-4 CDNA.
XX EXOSTOSIN-4; human; cytosolic; immunosuppressive; anti-HIV; Ext-2;
KW treatment; cancer; tumour; genetic disorder; AIDS; diagnostic; screening;
KM acquired immune deficiency syndrome; ss.
XX

OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 263..3022
FT /tag= a
FT /product= "Exostosin-4"
XX
XX W09925822-A1
XX PD 27-MAY-1999.
XX PE 14-NOV-1997; 97MO-CN00126.
XX PR 14-NOV-1997; 97MO-CN00126.
XX PA (UYHU-) UNITV-HUMAN-MEDICAL.
XX PL Deng H, Fan C, Liu C, Ruan Q, Xia J, Xu D.

DR WPI: 1999-347470/29.
DR P-PSDB: AAY08481.

PT New Exostosin-4 gene useful in the treatment of cancer, tumors,
genetic disorders and AIDS

PS Claim 2; Page 22-24; 33pp; English.

CC This invention describes a novel human Exostosin-4 polypeptide which has
CC cytosolic, immunosuppressive and anti-HIV activity. Exostosin-4 has
CC homology with Ext-2 and is therefore expected to have similar biological
CC activities. Exostosin-4 and its modulators can be used for treatment of
CC cancer, tumors, genetic disorders and AIDS (acquired Immune Deficiency
CC Syndrome), and in diagnostic/screening assays for such conditions.
XX

SO Sequence 5854 BP; 1204 A; 1611 C; 1611 G; 1428 T; 0 other;

Query Match 94.5%; Score 5833; DB 20; Length 5854;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 5836; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 332 GGCACCTTTTATCGTTTATAAAGTTAAGACATGTTTGGTCAACAGCCAGAACTT 391
|||||
Db 1 ggcaccttctacgttctgataaagattaaagacatgtctcttgcacagccagaaact 60
QY 392 AAAATCTGCTGAATAGAGGTGAGAGACCATTTTACGCTGACCTGAGAAATGAATGTT 451
|||||
Db 61 aaatctgctggaatagaggtgagagacccattcgcagctgaggaatgaatgct 120

QY 452 CATTTATTTGTCGCTTGTGCGAGACACCTTCTTCTGGAAGCTGTCAGTAA 511
|||||
Db 121 catttaattggtgctctgtctcgtggaagacacaaactctctggaagcgtgtaagaa 180
QY 512 ACAGAGATCGTTTGTGGAATAGAACCATGATTATGCGAGTACCCGACGATCTG 571
|||||
Db 181 acagagatcgttctgtggaatagaaacacaaatggtta tggcgagtgagaccgagcgtactg 240
QY 572 GGGGGCAGCTGACAGAGACTCATGACAGCTATACATGCTGCGGAATGGGGCGCGG 631
|||||
Db 241 gggggcagctgcagagacatcagacaggtacaaatgctcgtggaatgagagcgcgag 300
QY 632 GAACGAGCTGACAGCTGACAGCTGACAGCTGACAGCTGACAGCTGACAGCTGAC 691
|||||
Db 301 gaacgagctgacagctgacagctgacagctgacagctgacagctgacagctgacag 360
QY 692 CTTCACGCTCTTGTCAATCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 751
|||||
Db 361 cttaacgctcttctgacatctgctctctctctctctctctctctctctctctctctc 420
QY 752 TCTGATGAGCTGATGAGCTGATGAGCTGATGAGCTGATGAGCTGATGAGCTGAT 811
|||||
Db 421 tctgataagctgataagctgataagctgataagctgataagctgataagctgataagct 480
QY 812 GTCCGAGCTGAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 871
|||||
Db 481 gtccgagctgagacgctgctgctgctgctgctgctgctgctgctgctgctgctgct 540
QY 872 CCTGACGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 931
|||||
Db 541 cctgacgctgagagctgagagctgagagctgagagctgagagctgagagctgagagct 600
QY 932 GATGAGGCTGATGAGGCTGATGAGGCTGATGAGGCTGATGAGGCTGATGAGGCTGAT 991
|||||
Db 601 gatgagagctgataagagctgataagagctgataagagctgataagagctgataagag 660
QY 992 TGTATCAGCTGACAGCTGATGAGGCTGATGAGGCTGATGAGGCTGATGAGGCTGAT 1051
|||||
Db 661 tgtatcagctgacagctgacagctgacagctgacagctgacagctgacagctgacag 720
QY 1052 GTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1111
|||||
Db 721 gtccctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 780
QY 1112 CACTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1171
|||||
Db 781 cactgcgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 840
QY 1172 CTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1231
|||||
Db 841 ctcccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 900
QY 1232 GGTGACGAGCTTGTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1291
|||||
Db 901 ggtgacgagcttgtgagagctgagagctgagagctgagagctgagagctgagagctgag 960
QY 1292 CATGCGCTGCTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1351
|||||
Db 961 catgctgcttctacgtgataacagtgagagtgagagtgagagtgagagtgagagtgag 1020
QY 1352 TGTGAGCTGAGAGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1411
|||||
Db 1081 tgtgagctgagagagctgagagctgagagctgagagctgagagctgagagctgagagctgag 1140
QY 1472 TGGCGTGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1531
|||||
Db 1141 tggcggtgcataatgagagctgagagctgagagctgagagctgagagctgagagctgagag 1200
QY 1532 CTTGCTGCTATCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1591

Db 1201 cttggtcgtatcaaccgctggtccatctgctgtgagcccaacttcaatgaaatccacc 1260
OY 1592 ACAGGGGCGCGTGAACCGGAAATATCTTTCACCTTCAGGGCGAGAAATGAGTCTCT 1651
Db 1261 aacgggtgcggtgaaacggaataatctctcaccttccaggcgagagaatltgactctc 1320
OY 1652 GAAGCTAGCCCTCAGAGAGCCGCTCTCTTCGAAGAGGAAATGAGGGGACCCCTCCGC 1711
Db 1321 gaggctacgcttcagagagcccgctccttcgaagagaaatgagggcgaaacccctccgc 1380
OY 1712 CGACTACGATGACCGGATCATTCGCCACTGAAGCGGTGCAGAGACGAACTGATGA 1771
Db 1381 cgactacgatagcagatcatctgcacccctgaagcggtgcagagacaagctgtgata 1440
OY 1772 GGCCCGGTGGGAAATACCTGGAAAAACCAAGCCCAACCAAGCTCCGATGAGTGGC 1831
Db 1441 ggtccgggtggaattcaacttgcanaaaacagcccaaccacccctgcacactgagtggc 1500
OY 1832 ACTGTGTGAGAGCGGAGGAGCGGCTTGGAATTTGCTGAAGCTCTCCACCTTGGCCCTAT 1891
Db 1501 actgtgtgagagcgagagagacgcttgaattctgcgaagcctcccaacttcgacctcat 1560
OY 1892 CATTAACCCCGGAGACCTCGCTTGTTATTTCTCTGGGTGCAACACGGCTCTTGA 1951
Db 1561 cattaacccccggggaacccctgctgttatttccctctgggtgtgacaacccgctctcga 1620
OY 1952 AGCCCTGGAAGTGGGTGCGCTCCGCTGCTGGGGAGGAGGTCCAGCTTCCCTACCA 2011
Db 1621 agccctgaaagtcggtccgtccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1680
OY 2012 GGACATGCTGAGTGAAGGAGGAGCGCCCTGTGTGTGCAAGACCTCGTGAACGAGGT 2071
Db 1681 ggaatctgtcagatgtgagagcggtccctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1740
OY 2072 TCATTTCCTGCTCAAAAGCTCTCCGATGATGACCTCTGCTAGTGAAGCGGCAAGCGC 2131
Db 1741 tcaattctctgcagaagcctctcgtatgatacttctgtgtgtgtgtgtgtgtgtgtgtgtgt 1800
OY 2132 CTTCCTCTGGGAGACTTACTTCTCCACGCTGAGAGTATTTTATATACCGTCTGCTGAT 2191
Db 1801 cttctctcgtgagacttacttctcctcaactgcagatattttaaactgtgtgtgtgtgtgtgt 1860
OY 2192 GATTAGACTGCGATCGAGATCCAGCCGCTCCATCCGAGAGAGCGGAGCTGAGAT 2251
Db 1861 gattagagctcgtatccagatccagccgctcccaatccggaagagtggtgcagctgagat 1920
OY 2252 CCCCCACCGTTCAGGAGGAGGCGGTGGAAGTGAACCCCAACATGGCTGACAAAGGGGACT 2311
Db 1921 cccccacggttcaagcaagcggtgtgtaactgaccccaacaatgtgtgtgtgtgtgtgtgtgt 1980
OY 2312 GGACCTGGGCGGAGGAGGAGCGGCGCTAGCGCTACCCGACCGAATACCTCCGCAATTT 2371
Db 1981 ggaacttggtgcagatgagagagcgccctacgcctcacacagatactcccgcaattt 2040
OY 2372 CACTCTGACTGTACTGATTTTACCGAGCTGGAAGTGTCTCCAGGGCTTTTCACTCT 2431
Db 2041 caactctgactgactgacttacttccagctgtgaaactgtgtgtgtgtgtgtgtgtgtgtgtgt 2100
OY 2432 TTTTCCCCACACTCCCTTTTGAACCTGTGTGCTCCTCAGAGGCAAAATTTTGAGGCTCAG 2491
Db 2101 ttcccccaactcctcttgacctgt 2160
OY 2492 GACTGCTTTCGCGCTATTTGTGTGTGAGCTGGGGGTTCGGCAAGAAATTTTACGAGAC 2551
Db 2161 gactgtcttcgacctactgt 2220
OY 2552 GCTTGGAGCAATGTTCCCGAGAGAGCTCAGGCTGTGATGTTGACTTATGAGCGGGA 2611
Db 2221 gcttgagagaaatgttcccgagagcagttcaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2280
OY 2612 GGAAGTGTATGAACTCTTTAGAGAGGCTGAATGGCTCTCCCTTACTGAACAAGTGTGT 2671
Db 2281 ggaagtgtctatgaactctttagagagtgtaagtggtcccttacttgaacaagtgct 2340
OY 2672 GGTGTGTGGAATTTCTCCAAAGCTGCCATCAGAGGACCTTCTGTGGCGGACATTTGGCGG 2731
Db 2341 ggtgtgtgtgaattcttcccaagctgtcccaagctcgaagagagaccttctgtgtgtgtgtgtgt 2400
OY 2732 TCCATCATGTGTGTGCTCCGATCTGAGAGAACAGTTTGAACCAACGATTTTACCTTGAA 2791
Db 2401 ccccatcattgt 2460
OY 2792 TGAATTTGAGACGAGAGGACCATCTGTGCATTTGATGAGTGAAGTGTCACTCCGCAATGACA 2851
Db 2461 tgaattgtgaacagagagccatctgtccatltgatgagatgtgtgtgtgtgtgtgtgtgtgtgt 2520
OY 2852 AATCATGTTTGGGTTCGCGGTGTGAGAGAACCTCGGAGCCCAATGCTGGGCTTCCCTGG 2911
Db 2521 aatcatgt 2580
OY 2912 CCGTTACACGATGGGACATCCCATCATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2971
Db 2581 ccgttacccagcatgagacatcccatcagctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2640
OY 2972 TGAGCTGTCCATGT 3031
Db 2641 tgaagctgtccatgt 2700
OY 3032 TTCTTATGTATGCCCAAGGACATCCGGGACATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3091
Db 2701 ttcttatgt 2760
OY 3092 CATTCGCATGAATCTCTTGT 3151
Db 2761 catgtccatgaactctcttctgttccacatcaactgtgaagcccccaatcaagtgtaactc 2820
OY 3152 ACGGTGACATTTCCGATCCCAAGATGCGCTCAGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 3211
Db 2821 aagtgagacatctcogagtcccaagtggtcccaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2880
OY 3212 CCAGGAGCGGACAAAGTCAATCAACTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3271
Db 2881 ccaagagcggaacaagtgatcaacttctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2940
OY 3272 GTACACGAGTTCAGGAGT 3331
Db 2941 gtacaagagtgctgt 3000
OY 3332 CAAGTGTCTCAAGTTCATCTAGGGGAGCGAGCGTGTGGGAGAGATGAGAGAGG 3391
Db 3001 caagtgtctcaagtcatctcagtggt 3060
OY 3392 AGGAAGATGCTCCCAAGTTCCTAGGCATTTGAGGACCTTGGGACATCTGCTGTGTGG 3451
Db 3061 aggaagatgt 3120
OY 3452 TGGCCGAGAGCTCTGCTGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3511
Db 3121 tggccagagctctgt 3180
OY 3512 TCTTGAAGTGAAGCAGCAGCTGGGCTGTGAGCGCTGTGGGAGTGTCCCGGGGTTCGCCAGAC 3571
Db 3181 tcttgaagtgtaagccacactgt 3240
OY 3572 AGGCACTGATGATAGCTTACACTGAGAGCTGTGGAGCTGTGAGAGTCACTCAACAC 3631
Db 3241 agggcaactgtatagcttaccactgt 3300
OY 3632 GTTCGTACGCCCGAGAGAGCTGT 3691
Db 3301 gtctgtacgcccagagacagctgt 3360
OY 3692 ATTTAAAGAGAGAGTTCAGATTTGCAATTCGAAGCTTATATATATATGTGTGTGTGTGTGT 3751
Db 3361 atttaaaagagaaagttcagatctgtccatcaagcttatattatataatgtgtgtgtgtgtgtgtgt 3420

OY 3752 ATATAATACATGCACACTTGGCATATATATATTTTGGCTGGGGAGCTGTGAGTTT 3811
|||||
Db 3421 aataaataacatgcacacacttgcatatacatattcttgccgcyggggagtgtagcttc 3480
OY 3812 TGCCCTTCTAAGGAGAGGACCGGCCAGCTCCCTTTGTTCTGTATCTTGGCGGAATGGGT 3871
|||||
Db 3481 tgcctcttctaaggaggagacgycgcaaggctcccttgctcgtatcttcgycggaatgagtt 3540
OY 3872 CCTGGCCTTGATGTCATGCGCTTATCTTAAAGATCATCTCCCATCTCCAGCCAGCCATC 3931
|||||
Db 3541 ccctggccttggtctaacctggctataccttaagaatacctcccatcctcccaagcgccatc 3600
OY 3932 TGTGTGACCAACCAAGAAAGGATGAATTGGCCCTCTTGGCGGCGCTGGACAGGCTCTCT 3991
3601 tgtgtgcagcaacagaaagagatgaacttggccctcttgggccgagacaagttcctc 3660
OY 3992 TCCCTTACCCCTTCTGTTGGCCAGTACAGCAACTGTATCATCTCTTCCCACTGAATC 4051
|||||
Db 3661 tccctaacctctcgtctgcagtcagaaacctgaactcaatctctctcccaagtgaatc 3720
OY 4052 CCTGGAGCGCCTGACCCCTGGTGGCTGTTCAGCTTCTGCTGGCGGCGCAGGATTTT 4111
|||||
Db 3721 ccctggagcgccctgagccctggctggctgtctcagcttccgtgcttgaggccagatcttc 3780
OY 4112 TGAGGATTTATCTTTAAGGCCAGGCTTGCCCTCGTACTTATCCCTGCTCTCCATTTCTCT 4171
3781 tgaagatttacttcttaagccagsgctgctccgtaactatcctccgtccatcttccct 3840
OY 4172 CTCTTTTGAGAGAGATATGAGAGCAAAAGAGTGAAGAAAGATAGGGCTGAAGACGCCAC 4231
3841 ctcttcttgaaggagaaatgaaggaaacaaagagtgaaagaaatgaagggtgaaagcgccac 3900
OY 4232 TCCAGATGCGCTCTTGTATCTGCTCTCTGTTGAAACACACCTGCTGGGCTCAGG 4291
3901 tcccaagatgctctctctctatctctctctctctctctctctgtaaaacaacgtgctggctccag 3960
OY 4292 CGTTTCTGAAGTGCCTCTTCTTGTGATTTGGACAGAGATCAGACCGTGCACATCTGCTGT 4351
3961 cgtctctgaagtgctc 4020
OY 4352 GGTGTGAAGTGGTTTGAAGGTCAGCTCCCTCCCTAGTGTGAAGCAAGCCAGTGCCTT 4411
4021 ggtctgaagtggtc 4080
OY 4412 CGAGGAACCCACCCCGCTGGCGCGGGAAGTTTACAGCAAGCGCCTGCTTGGATATT 4471
4081 cgaggaaaccaccccgctggcgagggaagttttacagcaaggcgctctctgggataatc 4140
OY 4472 CCTTGGTGAATTTACCTTCCCGCGCTCTGTCTGAGGCCCATCTGTGTATCTGTG 4531
4141 ccttggtgaaatltcacacttcccccgcctctgcttgagaccacatccgtgtatctctg 4200
OY 4532 GTTTTGGACCCCTAATGTACGTTGGCTGAGAGTCGCCGAGATTGGATGTGCTAG 4591
4201 gttcttggaacccctcaagtgcaagcttggtctgtagaacctcccgaaggtcttgtaatgtctag 4260
OY 4592 AACAAATGGAGGCTGTGATTTGTGTGTGAAGCTACATCCAGCTTGGAAATCTAAGGGC 4651
4261 aacaattggagggtgtgattgtgtgtglaagctacatccagctctgaaatctaaagggc 4320
OY 4652 ATTACAAACCCGAGTTACCATTTCCACATCCCTGCTTAGATTTCTGTTCCCTGGGCTGAA 4711
4321 attacaacccgagltcacacttcccatctccctctctaggaatltcgttcccttgggctgaa 4380
OY 4712 ACTGAATTAAGCTAATTTTGGGTCAGGTCAGGTCAGTGAAGGGAACCTAAGAGGCTGTAG 4771
4381 actgaataaagctaattcttctgggtcatgtgtgaagtaagggaacctaaaggaggtgtgag 4440
OY 4772 TGGCAATTTGTACGGGATTTAGCCCATGAAGTGTCTTGAACCCCTACTTCTTGAAGTGG 4831
4441 tggcaatttgcagggaattttagcccatgtagctgttctctgaaacctacttcttgaaagtg 4500

OY 4832 AGTTGACTCTGGAAGTTTCTAGCAACTGAAACAAACACTAGGTTTCTCTGCTCATGCA 4891
|||||
Db 4501 agttgactctggaagtttcttctagcaactgaaacaaagctcaagttcttctctgttctagca 4560
OY 4892 CATGCCTTAAGCCAGTCCGCTTCCCTAGACCTTGGCATCTGCTGCTTCAATTTCTTGG 4951
4561 catgctttaaagccaagttccgtcttccctaaagaccttggaacctgtgtcttcaatctcttg 4620
OY 4952 AATGACTTCTCTGTGACACGTCGCTGTACACGTCGGGTCTTTCATGACTGTTTTGAG 5011
4621 aataagttccctctctctgaccgcgcctgtacaacgttggtccctctcaagtactgtttgag 4680
OY 5012 CTGGCTCTTTTGTGTAGCTCCACCCACCTGATAGGCTATGCTGCTTAAAGGACTCT 5071
4681 ctggcctcttcttgtagcttcccaaccactgttagggcttagctcggtcttaaggaaacttc 4740
OY 5072 CCCCATTTGGAAACCGGACCGCGCGCGGACAGGATGTTTCCAAAGGTTCCCGGCC 5131
4741 ccccatctgcaaaccggaacctggcgccgcgcgaagacctgtgtctcaaaagttcccgcc 4800
OY 5132 CCAACCCCAACATCAGCCTGTAGCTCCCTGCTGAGGACAGTGTGTTATGTTCCACAG 5191
4801 ccaaccccaacatcagcctgttagcttccctctgtgaggaagtggtatgttctccagcag 4860
OY 5192 TGGGGTCAGACGCCCTTCTCTAGAACTTTCTAGTCCCTCTACCTGACTCTGACTTG 5251
4861 tgggggtcagacgcgccttctccagaaacttcttagtttgctcctccagctctctgacttc 4920
OY 5252 TATTCCTTTTAGCGATGAGCGCTTCTCCCTGGGGAGCAAAAGATGTTGTGGGCGT 5311
4921 tatctctttagcagtagcctctctctccctcgcgggagcaaaagatgtgtgtggcgct 4980
OY 5312 ATATTGTGCGTCAATTTCAATCTGTTCTTTTAAATGTGAAGAACTCACATCTGACTTC 5371
4981 atattgtgcgtcatttcaattcattctgttctttaaigtgggaactcaatacctgacttc 5040
OY 5372 AGTGGACTGGGTAGGCCGGGGCCGCTGTGTGTGGAGACCCCTTTAGCGGACTAGT 5431
5041 agtggactgggtgagccggggccgtctgtgtgtggagacccctttagggggactagtc 5100
OY 5432 GAGCTGGGCGGCTGTGTGTGTGAGGACAGGCGCTTCCCTTATGATGAGCAAGTGTGTC 5491
5101 gagctggggcgctgtgtgtgtgtgagacagggccttccctctttagtgtagccaggtctgctc 5160
OY 5492 GGGCCCGAATGTCATGTTGTGATCTAAGAAAGGCTAGTGTGTGACACCAAAACATGC 5551
5161 gggcccgaaatgtcacctgtgatactaaaggaaggctgagtgtctgacacaaacaatgc 5220
OY 5552 CGCAGGAGGCGTGTGGTGGCGGCTTCCCAACAAGAGACGCCCTGTGACCTGAAAG 5611
5221 cgcaggagggtctgtgtgtccggtgtcttcaacaagaagacgcttctctgaccttgaaag 5280
OY 5612 GAACACTGCGTTTGAAGACATGACAGACGCTGTAGGGGACGCGCTCTCAGCGAAGG 5671
5281 gaacactgcttgaaagacatgcgaagcgtctcttgaaagggcaagcctctccagaggaag 5340
OY 5672 CAGCAAGTGGCCACACTGTCTAGTGTGAGTGTCTTACACACGGAAGAACGCGCGACC 5731
5341 cagcaagtggtgcacagtgtaactgttcaagtgcttcttccacaaaggaaagccgcagacc 5400
OY 5732 TGTGACTGCGTTTGAATGGGAAAGCGGCGACACAGACCCGGGCTCTGTTGGCTGTGT 5791
5401 tgtgactgcgttgaaatgaggaaacggcgcaacagaccgggctctctctgtgcgtctgtc 5460
OY 5792 GGGCGCGCCCTGAGCACCTTGTCTGAGCTGTGACAGGTGCAAGGACGCGCTGTCTGGG 5851
5461 gggcgccctctggcacactctgtctgtctcgcaagggtgacaggaagcgtctctctcgg 5520
OY 5852 TGGCGGCTTGTGCTCCGGTTTGGGCTGTCTTAACATAACACCGTCCAGGCGCTTGCA 5911
5521 tggcggcttgtgctctcgtctcgttcttggtcttcttcaacaacacgctccaggtctcgca 5580
OY 5912 GGCCACTGTAGGCGCTGGCTCCCTCGGGGACATGCTCCTCCGTTGTGACTGTGTGCTCAGGCC 5971

Db 5581 ggcactgtgagcgtgcctcctggcagtgctctccgtgagctgctcagcc 5640
 QY 5972 AGGCGTCACACGCTGGGCTCTCTCCGGAAGATGGATCTTTCTGGAGCTGCCGGA 6031
 Db 5641 agggcaccacagctgggtctctccggaagatggtatcttctcggagctgcgcgga 5700
 QY 6032 CAGAGTGGGAGAGCTCTAGTTTGGGGGAACTTGATATCATGCCACGTCATCCA 6091
 Db 5701 cagagtgaggagctcctagtcttggtgggaaagcttgatatacaccagcgcacaca 5760
 QY 6092 CCCACCCCTTTTGTGTACGACGACCAATGCTTACATGATTTTGTAAAAATATA 6151
 Db 5761 cccacccttctgcacagcagacaatgctctacattggtatcttctgaaaaataa 5820
 QY 6152 AATAATGAGACTTAACTC 6172
 Db 5821 aataatgagacttaactc 5841

RESULT 3

ID AAA12733 standard; cDNA; 3479 BP.
 XX AAA12733;
 XX

DT 25-JUL-2000 (first entry)

DE cDNA encoding mouse TRFX.

XX Tumour necrosis factor receptor-associated Factor: TRAF;

KW TRF-protein-interacting hereditary multiple extoses protein; TRFX;

KW signal modulator; tumour necrosis factor receptor;

KW CD40 mediated signal transduction; TRAF protein; cancer;

XX hereditary multiple extosis; autoimmune disease; ss.

XX Mus sp.

OS Location/Qualifiers

XX Key 458..3214

XX CDS /tag= a

XX /product= "tumour necrosis factor receptor-associated

XX 17-SEP-1998; 9805-0156191.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Sato T;

XX WPI: 2000-293180/25.

XX P-PDB; AAT84668.

XX New nucleic acid encoding Tumour necrosis factor Receptor-Associated

XX Factor (TRAF) protein-interacting hereditary multiple extoses (TRFX)

XX protein, useful in the diagnosing cancer -

XX Claim 20; Fig 7A-B; 161pp; English.

CC The present sequence encodes a murine tumour necrosis factor receptor-
 CC associated Factor (TRAF) protein-interacting hereditary multiple
 CC extoses (TRFX) protein. TRFX is a signal modulator which bridges
 CC between the tumour necrosis factor (TNF) receptor and CD40 mediated
 CC signal transduction. Anti-TRFX antibodies are useful for treating an
 CC abnormality in a patient by inhibiting binding of a TRFX protein and
 CC a TRAF protein (that is, TRAF 2, TRAF 3 or TRAF 5). The abnormality

CC is cancer, a hereditary multiple extosis or an autoimmune disease. The
 CC cancer is colon cancer, gastric cancer, human head and neck squamous
 CC cell carcinoma, prostate carcinoma, breast cancer, thyroid cancer,
 CC oesophageal cancer, lung cancer, colorectal cancer, ovarian cancer,
 CC papillary bladder cancer, osteosarcoma, chondrosarcoma, liposarcoma,
 CC giant cell tumour, Ewing sarcoma, or other malignant tumours.

XX Sequence 3479 BP; 762 A; 959 C; 951 G; 806 T; 1 other;

Query Match

Best Local Similarity 39.9%; Score 2459.6; DB 21; Length 3479;

Matches 2902; Conservative 0; Mismatches 505; Indels 40; Gaps 10;

QY 163 CTGTTCTAATCTTACTGATTAAGTGAATAGACATCATATTTGTGGTCAGCA 222
 Db 53 cgtttttaccagccgctgcatcact-tyaataagaagcatgcatatggtcgcgaca 111
 QY 223 AAACCAAGAAACAGACATATGCAATTTGAAAAAGCTGTGTGATTCAGAGGTGTTTTC 282
 Db 112 aagccaagggacaaaagctatgcccgttaaatgctcctgagtcacaggcctcttc 171
 QY 283 CTGGGTTTCATCAGTACGTAACCTCCCTTCATCTCAGCAAGATGTGGACCTTTTA 342
 Db 172 cggcctttagacacatgcatctccttctcaccatcagcaatggtgacccctct 231
 QY 343 TCGTTTGATTAAGATTAAGACATGTCTTTGTCTACAGCAGCACTTAAATCTGCTG 402
 Db 232 ctacttgatgag-----aaagctgatacttcagattgctc 268
 QY 403 GAATAGGCTGAGACCAATTCAGCTGACGTGAGAAATGAATGTTCAATTTATTTG 462
 Db 269 gactaaggttagaaccctgaalcgctg---tgaagaaatgaaattccatttacttg 324
 QY 463 GTGCTGTCTGGGGAGACACATACCTCTCTGGAAGCTGTGCAATGAAACAGACATGCT 522
 Db 325 gtccctgtgcaaggaagacatctatccctccagaactgtgtgtgaaa-agaagtg 383
 QY 523 TTTGTGAATAGCAACCCATGTTATGCGAGTAGCCAGCGATGATC---GGGGGCG 579
 Db 384 gtttgtagcagacagactatggtatagcgagcgatccgaagtatacgaagtggcaaga 443
 QY 580 GCTGACAGAGACTATGACAGGCTATACATGCTGCGGAATGGGGCGCGGGGAACGAG 639
 Db 444 ggcacagcgaactcatgacagcctataccatgctggaatggggagtggaagcggtg 503
 QY 640 GTCAACCTGCTGCTGGCGGTGTCACACCGCATCCCTACCTGCTGCTGCTACGCT 699
 Db 504 gtcaagcctgtagctgctgctgctgctcaatgcagcagctgacatggtggttcacgc 563
 QY 700 TCTTTGTATCCTGCTGCTCTCCCGCTCATCGCCACTATTAATCAACCACTGTGATG 759
 Db 564 tgtcatcatctctgcttcttctccctcatatgtaactatcaaccatcttgagc 623
 QY 760 AGGCTGATGAGGACGAGCGAATTTTGTGCTCCCGGTGGGAGACAGCTGTGCGAG 819
 Db 624 aggcagaagagctgagcagacatcttcgcccctcgctgagcagagctgagc 683
 QY 820 TGAAGCAGCTGCTGATCTGTCGCGCATCCGGCATCGGAGTGGTGAAGAGACTCTGAGC 879
 Db 684 taaagctgcttctgactctctgctgagctctgagctgtgagctgagcagagcttcaag 743
 QY 880 TGGAGGCCAAGCGCCAGAGCTGTAACAGCAGATGCGCAAGCTGAATCTGAAGTGAAG 939
 Db 744 tcgaagcagaagcagcagagctgaagaagcagatctgcaagctgaaactcaagatgag 803
 QY 940 CTGTGAAGAGAGCATTTGAGAACGCAAGCAGAGCACTGCTCAAGTCAAGATGTATCA 999
 Db 804 ccgttaagaagagcatagagatgacacagcagacactgctgacgtcacaagaatgcatca 863
 QY 1000 GCGAGACCGGCAATTCCTTCAAGAGAGCTCATGGCCCAAGAACGCAAGCTGTCCCTG 1059
 Db 864 gccagaagagcactctcaagaagctgagtcgccagaagaccagcccaactgtccctgc 923

[illegible]

QY	2140	GGAGACCTTACTCTCCACCTCGTAGCAATATTTTAAATACGCTGCTGGCTATGATTTAGGA	2139
Db	2001		2060
QY	2200	CTCGCATCCAGATATCCACACCCCTCCATCCCGGAAGAGCGCGAGCTGATATCCCCACAC	2259
Db	2061	ctcgaattccagatccccaactcgtctcccatccggaagagtagcggtctgaagttccccaatc	2120
QY	2260	GTTACGCGAAGCGGGCTGGAACTGACCCCAATGCGCTGACAAAGGGGACCTGGACCTGG	2319
Db	2121	gttcagcgaaagcaagctggaactgagccccaacaatgctgacaaatgaggacctggaacttgg	2180
QY	2320	GSCCAGTGGACACGAGCGCGCCCTACCGGCTACACCAGATTAAGTCTCGCAATTTCACTCGA	2379
Db	2181	ggccggtagagaacagaaaccacctatgctccaaactaaatlaacctcgaaatttcaactcga	2240
QY	2380	CTGTACAGACTTTTATACCGCAGCTGGAACTGTGCTCCAGCGGCTTTCCATCTTTTCCCCC	2439
Db	2241	ctgtcaacgaactglttaacgttggtcttggaactctgcccggagcglttccattcttttcccc	2300
QY	2440	ACACTCCCTTTGACCCTGTGTGGCCCTCAGAGGCCAAATCTTGGGCTTCAGGAGACTGGCT	2499
Db	2301	aaacacctcttgatccctgltgtgcccctctggagccaaattctctggctcaggaagctgagt	2360
QY	2500	TTCGGCGCTATTGGTGGTGGAGACTGGGGGTTCTGGCAAGAAATTTACAGCACCGCTTGGAG	2559
Db	2361	ttcggccgaactcgtgtggtcggtggtcttgggggtctctggaagaaattccaaaggcaagcgctcggag	2420
QY	2560	GCAATGTTCCCGCGAAGAGAGTTCAAGCGTGGTGGATTTACATTATGAGGGGAGGAAAGGC	2619
Db	2421	gcaatgtcccaagcgggaagagattcaacagtttggatgctgaactcaagagcggaaggaagtgc	2480
QY	2620	TTATGAACTCTTTAGAGAGGCTGAAATGGCCTTCCTTACCTGAACAAGGTCGTGGTGTGT	2679
Db	2481	tcaatgaactcccttggagagactcaacacggctcccccctaccctgaacaaagttagtgggtggt	2540
QY	2680	GGAATTTCTCCCAAGCTGGCATACAGAGACCTTTCTGTGGCTCGACATTGGGCTTCCCATCA	2739
Db	2541	ggaactctcccaagatgctccctcggggagaccttttggccagaaacttgggtgtccccaatca	2600
QY	2740	TGTGTGTCGGTACTAGAAAGAACTTTTGAACAAACCGAATTTCTAACCCCTGGAGATGAATTG	2799
Db	2601	tggtcgtccgtacgtcgtagaagaacagttcgaacaatcgttcttgcctctggaatgaaatctg	2660
QY	2800	AGACAGAGCGCATCTCTGTCCATTGTATGACGATGTCACTCGACCTCGGCCATGACAAATCATGT	2859
Db	2661	agacagagagccaactgctcatcagcagtgatgtctcacctctcgcaatgaatcaatcagt	2720
QY	2860	TTTGGGTTCCGGGTGTGTGGAGAGAGAGCTCGGGACGCGATGCTGGGCTTCCCTGGCGCTTACC	2919
Db	2721	ttgggtttctgggtgtgtgagagaagaaacagtgatcgcatgtgggtttccctctggccggtacc	2780
QY	2920	ACGCATGGGACATCCGCCATAGTCCCTGCTGAGGCTCTAACCTCAACATACCTGCTGAGCGTGT	2979
Db	2781	atgcgttgggaacttcccgacacagttccgtgcttcaaatlcaaacactccgttgaagctgtc	2840
QY	2980	CCATGCTGCTGACAGAGTCTGCGCTCTTTTCAACAATATATAGCCTCACTGTATTTTATG	3039
Db	2841	ccaatgtgtcgtacggtggtcgtcgtctcttcttcaagaatataatgacctatctctatgt	2900
QY	3040	TGATGCCCCAGCGCATCCGGGACATGTTGATGAATTAATCATCAACTGTAGGACATTTGCCA	3099
Db	2901	tgaatgccccagaagccaatccgggaacatctgtgagcagatatacatcaactgttgaagataatcg	2960
QY	3100	TGAATTTCTTGTGTCTCCCAATCACTCGGAAGCCCCCATCAAGTGAACCTTCACAGGTGGA	3159
Db	2961	tgaacttctctgtctccacaatcaacacggaaaccccccaataagttgacataaagtgga	3020
QY	3160	CATTTCCGATGGCCAGAGATGSCCTTCAGGCGCTGTCTCATGATGATGATCCCACTTCCACGAGC	3219
Db	3021	ctcttcgattgcccgaaggtgtccctctcagccctgttcccaatgaaatctcaatttcaagagc	3080
QY	3220	GGCACAATGTGATCAACTTCTTCTGTGAAGGTGTACGGCTACATGCCCCCTCTGTACACGC	3279

Db 3081 ggcacaaagtatcaacttttgcgaagtgtaagcgtatatacctctcttgtaaacac 3140
 QY 3280 AGTTCAAGGGGATTTCTGTGCTTTAAGACAGCCTGCCCATGACAAAGACCAAGTCT 3339
 Db 3141 agtctcgagtgagctctctctcaagaccgctgcctccatagcaagaaccaagtgtc 3200
 QY 3340 TCNAAGTTCAATCTAGGCGACGACGCTGGGGAGAGAGATGACAGAGGAGAAAGT 3399
 Db 3201 tcaagtctctcagagcctctcagctctcgaagagacaatgagcag-aggcgaaggaggagtc 3259
 QY 3400 GCGTCCCAAGGTTCTTAGGCAATGACGACCTTGGGACATCTGCTGTGGTGGCCAG 3459
 Db 3260 accctcaaggtcccaaggtgcgaaggtcccttgaggacatctg-tcggcgaaggcccaag 3318
 QY 3460 AGCTCTGCTGGAAGGGGACGAGAGAGAGTGAAGAAACCGCTGCCCTT-ATCTTGA 3518
 Db 3319 acccttgctggaagagagcagagaaagtgaaaggatagctgctctcatcttga 3378
 QY 3519 GTACGACACACTGGGCGCTGAGCGCTGGGAGAGTCCCGGGGTGCCCCACAGAGGAC 3578
 Db 3379 gtcaagcacaactggcctggagatcctctgacagaagactcagncgt--ctgcacaggccac 3436
 QY 3579 TGACTGATAGCTTACACTGAGAGACTGT 3605
 Db 3437 tgactgatagcgaacactgagagctgt 3463

RESULT 4

AAF26844 ID AAF26844 standard; cDNA: 3198 BP.

AAF26844;

05-APR-2001 (first entry)

Rat Reg-binding protein encoding cDNA SEQ ID NO:3.

Rat; Reg-binding protein; Reg receptor; Reg; DNA synthesis; apoptosis; cell proliferation; diabetes; antidiabetic; ss.

Rattus norvegicus.

Key Location/Qualifiers
 CDS 5..2764
 FT /*tag= a
 FT /product= "Reg-binding protein"

W020007192-A1.

21-DEC-2000.

09-JUN-2000; 2000MO-JP03764.

10-JUN-1999; 99UP-0164488.

(OKAM/) OKAMOTO H.

Okamoto H;

WPI: 2001-061871/07.

P-SDB: AAB51341.

Protein binding to pancreatic Reg protein and its encoding DNA, useful for screening candidate antidiabetic drugs -

Claim 1; Page 64-74; 11pp; Japanese.

The present invention describes DNA which encodes: (a) a protein of rat origin binding to Reg protein; or (b) a Reg-binding protein derived from the natural Reg sequence by addition, deletion and/or substitution of one or more amino acid residues. The present sequence encodes a rat Reg-binding protein from the present invention. Reg-binding proteins

CC have antidiabetic activity. Drug compositions comprising Reg-binding proteins and their genes are useful for the treatment and/or prevention of diabetes. In addition, they are useful in developing new treatments for diabetes.

Sequence 3198 BP; 723 A; 882 C; 850 G; 743 T; 0 other;

Query Match 38.0%; Score 2347.4; DB 22; Length 3198;
 Best Local Similarity 85.3%; Pred. No. 0;
 Matches 2721; Conservative 0; Mismatches 416; Indels 54; Gaps 7;

QY 590 ACTCATGACAGCGTATACCATGCTGGGAATGGGGCGGGGGAAGAGTACAGCTG 649
 Db 1 actcatgacagcgtataccatgttgagatgggagagggagagcagcgtgagcgtgcatcat 60
 QY 650 CATGGCGCGCTGGTCAACCGCATCCGCTCAACGTCGCTCAGCTTCCAGCTTTGTCT 709
 Db 61 tatgctgcgtggtcccaaccgcatccgctgagcctggtgagttcaacgctgtcatcat 120
 QY 710 CCTGCTCTCTTCCCGCTCATGCGCCACATTAATACCTCAACACTGTGATGAGCTGATGA 769
 Db 121 cctgctctcttcccccctcatgcccactatactcaacactcggatgagcagatga 180
 QY 770 GGCACGCAAGCGGATTTTGGTCCCGGGTGGGGAAGAGAGCTGCGAGGTGAACGACT 829
 Db 181 ggcgcaagcgagcttcttgcccggtggtgcaagagctctgtggaatgaacagct 240
 QY 830 GCTGATCTGTGCGCATCCGGAGTGGTGTGATGAAGCTCCGTCAGCTGAGGCCCA 889
 Db 241 cctgactcttctgctgagctcgagctcgagctgagcgaagactctcaagcagtagaagc 300
 QY 890 GCGCCAAAGCTGAACACAGCATCGCCAGCTGATGATGAATGACAGCTTGAAGAA 949
 Db 301 gcggcagagctggaacagcagatgccaagctaacctcaagatgaagctgtaagaa 360
 QY 950 GAGCATTTAGAAGCGCAAGAGAGAGCTGCTCAGCTCAAGATGTCATCAGCCAGACCA 1009
 Db 361 gagatagagaacgcgaagcagagcctgtcagctcaagaatgcatltaagcagaaga 420
 QY 1010 GCATTCTTACAAAGAGCTCATGGCCCAAGACCCCAAGCTGTCCCTGCCATCCGACT 1069
 Db 421 gcaactctacaagagagctgtagcgcgaacacacacacacacacacacacacacacacac 480
 QY 1070 GCTCCCAAGAGAGAGATGCGGCTCCCTCCCGCAAGAGCCACTGGGGCTCCGGCT 1129
 Db 481 gctcccgagaagatgtagcgtgcttccaccccccaagtaactcgtgggtgctgcgct 540
 QY 1130 ACACAATGCTTTATTTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1189
 Db 541 acacaactgcttgcattactctgcttgcctctgacgtctgcttctctctcgtcctc 600
 QY 1190 TGACAGTACACAGTTGCTTTGGCAGCTACCTGATGATGATGATGATGATGATGAT 1249
 Db 601 tgacagtgacagtttgccttgccttgccttgccttgccttgccttgccttgccttgcct 660
 QY 1250 GCGGACAGCAGAGCTTACGTTATGTTACAGAAATGACAGATGCGCTGCTTACGT 1309
 Db 661 ggcacagtagagccaaagcttattatgacagaaatgacacatgcgcctgctatgt 720
 QY 1310 GATACTAGTGGGAGATGAGAGAGCCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1369
 Db 721 ggtgtagtgaggagtagtaaaagccgctgctgctgctgctgctgctgctgctgctgct 780
 QY 1370 GTTGATTTCCCTCCACACTGGGAGAGAGTGAACAAACATGATCATCATCATGCTG 1429
 Db 781 gctgcatctcttgcacactcggagagagagagagagagagagagagagagagagagag 840
 QY 1430 AGTAACTGATGATGACAGAGACCTTCTATTAAGTCACTGCTGCTGCTGCTGCTGCTGCT 1489
 Db 841 ccggaagtcagacacacaaattactgtatacaatgcaagtcagtcagtcagtcagtcagtc 900
 QY 1490 CCAAGTCCACTTCTACACTGTCCAGTACAGACCTGCTGCTGCTGCTGCTGCTGCTGCT 1549

Db 901 cccgctacactctctatctgctgcccagctacagagcttgaccttgacttgctgctacacact 960
QY 1550 GGATCATGTCATGTCATGAGCCCAACTTCATGAAATCCACACAGAGTCCCGGTGAGCG 1609
Db 961 tctgcatatgcaatctctgaaacccaactctcatgtaaaatcccaacgagctgcagattaaagc 1020
QY 1610 GAAATATCTCTACCTTCCAGGGCGGAGAGATTGAGTCTCTAGGTCTAGCTTACAGGA 1669
Db 1021 gaaatctctctacactctctccaggtctgaaagattgagctctctaaagcttagaccttcagga 1080
QY 1670 GGGCCGCTCTTGGAGAGAGAAATGAGAGGGCGACCTCCCGCCGACCTACATGACCGGAT 1729
Db 1081 ggcgcgtctctctgagaaagaaatgagaggtgagccctccgagcagctatgatactgaat 1140
QY 1730 CATTCGCACCTTGAAGCGGTGACAGACAGACAGTGTGAGTCTGCTGAGAAATGCAC 1789
Db 1141 catctgcaacccctcaabgacgctacagaaacagcaagctcagatcaggtctgctgtaagaattac 1200
QY 1790 CTGCAAAAACCAACCCCAACCCAGCCCTGCCAGTGAAGTGGCACTGTGTGAGAGCGGGA 1849
Db 1201 ctgcaaaaaacagcaaaagccagctcgcctactgagctgagcaactgctgagagcgagga 1260
QY 1850 GGACCGCTTGGAAATGGTGAAGCTCTCCACCTTCGCCCTCATATTATTCACCCCGGGGACCC 1909
Db 1261 ggaacgctgtagagctactgaaagctctccacactcgcctacatcaatccacgagagcc 1320
QY 1910 TCCCTTGGTATTTCCTCTCTGGGTGTGCACACGCGCTTCTGGAAGCCCTGGAGAGCGGTGC 1969
Db 1321 gaaacgtctatctatctatctctgctgctgcaaaacagcctctctgaaagcctctgagagtgagagc 1380
QY 1970 GCTCCCGTGTGCTGGGGAGCAGGTCTACCTCCCTACAGACATGCTGACAGTGA 2029
Db 1381 ctgctgctctgctctctgagagcagctgacgtctccgctacacagactgctacagagaa 1440
QY 2030 CGAGGCGCGCTGTGTGTGCAAAAGCTTCGTATACCGAGTTCATTTCCGCTCAAG 2089
Db 1441 tggagccgcctcagctgagctgcaagcctcgtctacagaggttcaactctctgtaacagag 1500
QY 2090 CACTCTCGAATGATGACCTCTGCTATGAGAGCGGCAAGCGCGCTTCTGTGGAGACTTA 2149
Db 1501 tctgctgcaacagagatctgctgctgctagagcggcaagcgctctctctcggagagactta 1560
QY 2150 CTCTCTCAGTGTGACAGTATTTTAAATACCGTGTGCTATGATTAAGACTGCATCA 2209
Db 1561 ctctctcaacgctgacagatattcttaatacagctgctgcaagctagtagactcgaattca 1620
QY 2210 GATCCAGCGCGCTCCCATTCGGGGAAGGCGGCAAGCTGAATCCCCACCGTTTCAGGCA 2269
Db 1621 gatcccaagctgctcccactccggagagagctgagagctgagatccccaatgcttaagcaaa 1680
QY 2270 GGGGCTGGAATGATGACCCCAACATGGCTGACACAGGGGACTGGAGCCAGTGA 2329
Db 1681 ggcagctgtagtaacgaaacccaatgctgcaaatgagagacctggaacctgagggccggtaga 1740
QY 2330 GAGGAGCGCCCTTACCGCTACCCAGATACCTCCGCAATTTCACTGTGACTGTCACTGA 2389
Db 1741 gacagagcgcccatgctctcaactaaatactctgaattctacactgctgactgta 1800
QY 2390 CTTTATACCGAGCTGGAACGTGTGCTCCAGGGCTTTTCCATCTTTTCCCCACACTCCCTT 2449
Db 1801 ctcttcaacgagctgagaaactcgaaacccgagactcttcaactcttctcaacaacaacact 1860
QY 2450 TGACCTGTGTGTCCTCAGAGGCAAAATCTTGGGCTCAGAGGACTGGCTTTCGGGCTAT 2509
Db 1861 tgaacctgctgctgctccctctgagagcaaaatctccgggctcagggactgattctcgcccat 1920
QY 2510 TGGTGTGAGCTGTGGGGTCTTGCGAAGAAATTTCAAGCAGCGCTTGGAGCAATGTCTC 2569
Db 1921 cgggtggtggggcctgagggcctgagcaagagcttccagcgagcgcttggaggaatgctca 1980
QY 2570 CCGAGACAGATTCACGGTGTGTGATGTTGACTTATGACGGGAGAGAAATGCTTATGAACCT 2629
Db 1981 gcgagagcagctcaacagctgctgactcagacagcgagagagagctcactgaactc 2040
QY 2630 TTTAGAGAGGCTGATATGAGCCCTCCCTTACCTGAACAAAGTGTGTGTGTAATCTTCC 2689
Db 2041 cctgagagagctcaatagctctccctccactcagaaacaggtagctggtgctgagaaactctcc 2100
QY 2690 CAAGCTGCCATCAGAGGACCTTCTGTGCTGTGACATTGGCGTTCCATTCATGTGTGCTCG 2749
Db 2101 caagctgcccctcgagagacctcttgctgagacaaatctgctcccaatcagctgctcgcg 2160
QY 2750 TACTGGAAGAACAGTTTGAACAACGATTTCAACCTGTGAATGAATTTGAGACAGAGCG 2809
Db 2161 tactggaagaaacagctctgaaacaaatcgctctgcccctgaaatgaaagaaagagc 2220
QY 2810 CATCTGTCCATTTGATGACGATGCTCACCTCCGCCATGACGAAATCATGTTTGGTTCG 2869
Db 2221 aatatctgctcatctgataagcaatgcccaccccgcaatgaaatcaatctgctggcttcg 2280
QY 2870 GGTGTGAGAGAAAGCTCGGAGACCGCATCTGTGGCTTCTCCCTGGCGTTTACCACGATGGGA 2329
Db 2281 ggtgtgagagagcgctgatacgtcatctgagggtctcccgcggtacaaatgctgagga 2340
QY 2930 CATCCCCATCAGTCCGCGCTTACAACATCCCACTACTCTGTGAGCTGTGCAATGAGTCT 2389
Db 2341 catccctcaacagctccgctctcaacaaatccaaactccctgagagctgctacatgctc 2400
QY 2990 GACAGGTGCTGCTCTTCTTTCACAAGTATTAATGCTTACCTTATTTATGTGATGCCCA 3049
Db 2401 gacggggtgctgctctctcttcaacagatatacgcctactctgattctctatgtgctccca 2460
QY 3050 GGCCTATCCGGGACATGCTGTGATGATTAATCACTGTGAGGACATTTGCCATTTCTCT 3109
Db 2461 ggcctatccgagaaatggtgagatgataatcaactgagaaatcagcaatgcaactcct 2520
QY 3110 TGTCTCCCATCATCTGAGGAGCGCCCATCAAGGTGACCTCCAGGTTGACATTTCCGAT 3169
Db 2521 tgtctcccaatcaacaagaaagcccccaataaggtgaaatcagagtgagagctcttcgagctg 2580
QY 3170 CCCAGATGCCCTCAGGCGCCTGTCTCATGATGACTCCCACTCCACAGGCGGCAAGATG 3229
Db 2581 cccgggtgctcccaagcctctgtcccaagatgactctcaactctcaatgagagcaaggtg 2640
QY 3230 CATCACTTCTTGTGTGAAGTGTACGCTCATATGCCCCCTCTGTACACGACATTCAGGCT 3289
Db 2641 tatcaactcttctgtcaaggtgtaaggtctatgcccctctctgtacaaacaggtcttaggt 2700
QY 3290 GGAATCTGTGCTCTTCAAGACAGCGCTCCGCCATGACAAAGCAAGGCTTCAAGTTCAT 3349
Db 2701 ggaactctgctctctcaagaccgctgcaccaatgaaagaaacaaagtctcaagttcat 2760
QY 3350 CTAGGGGC-AGCGCACGCTGTGGGGAAGAGATGACGAGAGGAGAAATGCTGCCAA 3408
Db 2761 cttagagctctgcaagctctgagagaaagacagtgagagagtgagggagtgaaaccccaa 2820
QY 3409 GGTTCCT-AGGCATTTGACAGACCTTGGGCAACATCTGCTGTGGTGGCCACAGACCTCTG 3467
Db 2821 ggttcccaaggtctgaaaggtctcttgaggacatctggtggag--ggtcccaagggcctctg 2877
QY 3468 CTGGAAGGGGACAGGAGGAGGATGGAAGAAACGCTGCTTATCTGTAAGTCAGGCAC 3527
Db 2878 ctggaagaaag-agaagagaaagtagaaaggaatgagctctctcttaatttgaagltaaagcgc 2936
QY 3528 ACTGGGCTGTGAGAGCCCTGGGCGAGTCCCGGGGTTCCCAACACAGGAGGACTGAGTATA 3587
Db 2937 actgggctctgaaatcccgctcag-----cagactcaagtgagccagcgtactag 2982
QY 3588 GCTTACACTAGACACTGTGGCGACTGTGACAGATCACTCACACCGCTTGTGTAACGCCAAGA 3647
Db 2983 gcaaacactgagact-----gtcacaagagcccgga 3014
QY 3648 CAGCTGCTGTGTGTTTTCATTTCAATTAACAACATATATGATTAATTAAGAGAGAAAG 3707
Db 3015 cagctgctcccggtcttctaattcaagaagaacattactaattc-----aaagaaagag 3068

Db 1981 gcgagagcagctcaacagctgctgactcagacagcgagagagagctcactgaactc 2040
QY 2630 TTTAGAGAGGCTGATATGAGCCCTCCCTTACCTGAACAAAGTGTGTGTGTAATCTTCC 2689
Db 2041 cctgagagagctcaatagctctccctccactcagaaacaggtagctggtgctgagaaactctcc 2100
QY 2690 CAAGCTGCCATCAGAGGACCTTCTGTGCTGTGACATTGGCGTTCCATTCATGTGTGCTCG 2749
Db 2101 caagctgcccctcgagagacctcttgctgagacaaatctgctcccaatcagctgctcgcg 2160
QY 2750 TACTGGAAGAACAGTTTGAACAACGATTTCAACCTGTGAATGAATTTGAGACAGAGCG 2809
Db 2161 tactggaagaaacagctctgaaacaaatcgctctgcccctgaaatgaaagaaagagc 2220
QY 2810 CATCTGTCCATTTGATGACGATGCTCACCTCCGCCATGACGAAATCATGTTTGGTTCG 2869
Db 2221 aatatctgctcatctgataagcaatgcccaccccgcaatgaaatcaatctgctggcttcg 2280
QY 2870 GGTGTGAGAGAAAGCTCGGAGACCGCATCTGTGGCTTCTCCCTGGCGTTTACCACGATGGGA 2329
Db 2281 ggtgtgagagagcgctgatacgtcatctgagggtctcccgcggtacaaatgctgagga 2340
QY 2930 CATCCCCATCAGTCCGCGCTTACAACATCCCACTACTCTGTGAGCTGTGCAATGAGTCT 2389
Db 2341 catccctcaacagctccgctctcaacaaatccaaactccctgagagctgctacatgctc 2400
QY 2990 GACAGGTGCTGCTCTTCTTTCACAAGTATTAATGCTTACCTTATTTATGTGATGCCCA 3049
Db 2401 gacggggtgctgctctcttcaacagatatacgcctactctgattctctatgtgctccca 2460
QY 3050 GGCCTATCCGGGACATGCTGTGATGATTAATCACTGTGAGGACATTTGCCATTTCTCT 3109
Db 2461 ggcctatccgagaaatggtgagatgataatcaactgagaaatcagcaatgcaactcct 2520
QY 3110 TGTCTCCCATCATCTGAGGAGCGCCCATCAAGGTGACCTCCAGGTTGACATTTCCGAT 3169
Db 2521 tgtctcccaatcaacaagaaagcccccaataaggtgaaatcagagtgagagctcttcgagctg 2580
QY 3170 CCCAGATGCCCTCAGGCGCCTGTCTCATGATGACTCCCACTCCACAGGCGGCAAGATG 3229
Db 2581 cccgggtgctcccaagcctctgtcccaagatgactctcaactctcaatgagagcaaggtg 2640
QY 3230 CATCACTTCTTGTGTGAAGTGTACGCTCATATGCCCCCTCTGTACACGACATTCAGGCT 3289
Db 2641 tatcaactcttctgtcaaggtgtaaggtctatgcccctctctgtacaaacaggtcttaggt 2700
QY 3290 GGAATCTGTGCTCTTCAAGACAGCGCTCCGCCATGACAAAGCAAGGCTTCAAGTTCAT 3349
Db 2701 ggaactctgctctctcaagaccgctgcaccaatgaaagaaacaaagtctcaagttcat 2760
QY 3350 CTAGGGGC-AGCGCACGCTGTGGGGAAGAGATGACGAGAGGAGAAATGCTGCCAA 3408
Db 2761 cttagagctctgcaagctctgagagaaagacagtgagagagtgagggagtgaaaccccaa 2820
QY 3409 GGTTCCT-AGGCATTTGACAGACCTTGGGCAACATCTGCTGTGGTGGCCACAGACCTCTG 3467
Db 2821 ggttcccaaggtctgaaaggtctcttgaggacatctggtggag--ggtcccaagggcctctg 2877
QY 3468 CTGGAAGGGGACAGGAGGAGGATGGAAGAAACGCTGCTTATCTGTAAGTCAGGCAC 3527
Db 2878 ctggaagaaag-agaagagaaagtagaaaggaatgagctctctcttaatttgaagltaaagcgc 2936
QY 3528 ACTGGGCTGTGAGAGCCCTGGGCGAGTCCCGGGGTTCCCAACACAGGAGGACTGAGTATA 3587
Db 2937 actgggctctgaaatcccgctcag-----cagactcaagtgagccagcgtactag 2982
QY 3588 GCTTACACTAGACACTGTGGCGACTGTGACAGATCACTCACACCGCTTGTGTAACGCCAAGA 3647
Db 2983 gcaaacactgagact-----gtcacaagagcccgga 3014
QY 3648 CAGCTGCTGTGTGTTTTCATTTCAATTAACAACATATATGATTAATTAAGAGAGAAAG 3707
Db 3015 cagctgctcccggtcttctaattcaagaagaacattactaattc-----aaagaaagag 3068

Query Match 9.7%; Score 596.4; DB 22; Length 598;
 Best Local Similarity 9.8%; Pred. No. 2.4e-146;
 Matches 597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3696 AAAAAGAGAAAGTTTCAGATTTTGCATTCAGGCTTATTTATATATATGTGTGTATAT 3755
 1 aaaaagaaagatttcagatttcgcatcgaagcttatatatatgtgtgtat 60

QY 3756 AATTACATGCACACACTGTGATATATATTTTGGCGGGGAGTGAAGTTTGGC 3815
 61 aattacatgcacacactgtgatatatatatttggcgggaggtgtgtgttggc 120

QY 3816 TTTCTAAGGAGGAGCGCGCAGGCTCTTTGTTGTTATCTGGCGAGATGGTCTG 3875
 121 ttcttaaggagagcgcgagctcttctgtctgtatctcggagatgtgttccg 180

3876 GCCTTGTTGCTACGCTTATCTTTAAAGATCATCTCCATCTCCAGCGCATCTGTG 3935
 181 gctctgttgcactgtctatctcttaaaagatcattccatcccccagcgcatctgtg 240

QY 3936 TGCAGCAACAGAAAGGATGAATGGCCCTCTGCGGGGCTGAGCAAGCTCTTCT 3995
 241 tgcagcaaacagaaaggatgaatggccctctgcggggctgagcaagctcttcc 300

QY 3996 TACCTTTCTGTTGCAAGTACGCAACCTGTACTACATTTCTTCCAGTGAATCCCTG 4055
 301 taccttctgttgcagctgacgaacctgtactcaatctctctccagtgatccctg 360

QY 4056 GGAGCGGCTACCTGCTGGGCTGTTACGTTCTGCTGCTGGGCGCAGGATTTTGGAG 4115
 361 ggagcggctacctgctgggctgttaccgttctgctgctgggcgagcaatcttga 420

QY 4116 GATTATCTTAAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4175
 421 gattatctttaaagcggctgctgctgctgctgctgctgctgctgctgctgctg 480

QY 4176 TTTGAGAGAGATGAGAGCAAAAGAGTGAAGAAATAGGGCTGGAAGCGCCACTGCC 4235
 481 ttgagagagatgagagcaaaagagtgaagaagaataggggctggaagcgccactccc 540

QY 4236 AGATGGCTCTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4293
 541 agatggctcttattctgctgctgctgctgctgctgctgctgctgctgctgctg 598

ABL26683
 JLT 8
 ABL26683 standard; DNA; 3595 BP.

AC ABL26683;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 31522.
 XX
 KW Drosophila: developmental biology; cell signalling; insecticide;
 KM pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001MO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EM;

XX MPI; 2001-656860/75.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 31522; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB101840-AB16175), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB16175-AB16175).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pot_sequences.
 XX
 SO Sequence 3595 BP; 928 A; 897 C; 884 G; 886 T; 0 other;

Query Match 8.8%; Score 545.6; DB 23; Length 3595;
 Best Local Similarity 60.9%; Pred. No. 1.6e-132;
 Matches 944; Conservative 0; Mismatches 594; Indels 12; Gaps 3;

QY 1818 CCGAGTACGAGGAGGCTGTGTGAGAGCGGAGACCGCTTGGATTTGCTGAGCTCTCC 1877
 1818 ccgagtacgagagggctgtgtgagagcgagacccgcttggatTTGCTGAGCTCTCC 1877

QY 1878 ACCCTGCGCTCATATTTACCGGCGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1937
 1878 accctgCGCTCATATTTACCGGCGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 1937

QY 1938 ACACGCTCTTCAAGCCCTGGAAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1994
 1938 acacgctctTCAAGCCCTGGAAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1994

QY 1995 GTCACGCTCTTCAAGCCCTGGAAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2054
 1995 gtcacgctctTCAAGCCCTGGAAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2054

QY 2055 CTTGCTGTTTACCGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2114
 2055 cttgctgTTTACCGAGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2114

QY 2115 ATGAGGCGGCAAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2174
 2115 atgagGCGGCAAGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2174

QY 2175 AATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2234
 2175 aatACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2234

QY 2235 GAGCGGCGGCAAGCTGAGTCCCGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2294
 2235 gagCGGCGGCAAGCTGAGTCCCGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2294

QY 2295 GCTGACACGCGGCAAGCTGAGTCCCGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2354
 2295 gctGACACGCGGCAAGCTGAGTCCCGCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2354

QY 2355 AGATACCTCCGCAATTTTACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2414
 2355 agatACCTCCGCAATTTTACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2414

QY 2415 CGAGGCGCTTTCATCTTTTCCCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2474
 2415 cgagGCGCTTTCATCTTTTCCCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2474

QY 2475 AAATTTCTGGGCTCAGGAGCTGCTTCCGCTTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2534
 2475 aaatTTCTGGGCTCAGGAGCTGCTTCCGCTTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2534

Db 239 AGGTGCATCTGAGAGAGCAAGCCCTTGAGGTTCACTCTTTCAAGAAGTCGTGCTGAGG 180

Qy 73 TGTAAATGCTACACAGTCAGAGGAAGGAGGCTCTGAAACACATGCGCTGATTGTTGGC 132

Db 179 TGTAAATGCTACACAGTCAGAGGAAGGAGGCTCTGAAACACATGCGCTGATTGTTGGC 120

Qy 133 AAAGGCATCATTAAGAAAGCTGGCAATTTATTTCTGTTCTAACTATTACTGTTAACTGTGA 192

Db 119 AAAGGCATCATTAAGAAAGCTGGCAATTTATTTCTGTTCTAACTATTACTGTTAACTGTGA 60

Qy 193 ATAGACACTATGCATATTTGTTGCTGAGCAAAACCAAGA 232

Db 59 ATAGACACTATGCATATTTGTTGCTGAGCAAAACCAAGA 20

RESULT 13

AA21849

AA21849 standard; DNA: 3008 BP.

AA21849:

27-MAR-2001 (first entry)

Human breast and ovarian cancer associated antigen gene SEQ ID 236.

Human breast and ovarian cancer; cytosolic; immunosuppressive; neurotropic; neuropsychiatric; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer; vulnereary; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiac; immune disorder; Addison's disease; allergy; autoimmune haemolytic anaemia; autoimmune thyroiditis; diabetes mellitus; Crohn's disease; multiple sclerosis; rheumatoid arthritis; ulcerative colitis; cardiovascular disorder; wound healing; neurological disease; ds.

OS Homo sapiens.

XX

PN WO20005173-A1.

XX

PD 21-SEP-2000.

XX

PE 08-MAR-2000; 2000MO-US05681.

XX

PR 12-MAR-1999; 99US-0124270.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

RA Rosen CA, Ruben SM;

XX

DR WPI: 2000-611515/58.

XX

P-PSDB: AAB58946.

XX

PT New human breast and ovarian cancer associated gene sequences and the

PT polypeptides encoded by these genes, useful in the prevention,

PT treatment and diagnosis of cancer, immune disorders, cardiovascular

PT disorders and neurological diseases -

XX

PS Claim 1; Page 660-661, 129pp; English.

XX

CC Sequences AA21614 - AA22201 represent DNA sequences encoding human

CC proteins AAB58711 - AAB59128. The DNA and protein sequences are

CC associated with breast and ovarian cancer. Included in the invention are

CC sequences AA22032 - AA22040 and AAB59129 which are used in the

CC isolation and characterisation of the DNA and protein sequences of the

CC invention. The breast and ovarian cancer associated DNA, protein, agonist

CC or antagonist sequences exhibit cytostatic; immunosuppressive;

CC neurotropic; neuropsychiatric; antiviral; antiallergic; hepatotropic;

CC antidiabetic; antiinflammatory; anticancer; vulnereary; anticonvulsant;

CC antibacterial; antifungal; antiparasitic and cardiac activity. The

CC polynucleotide and protein sequences are used in the diagnosis of cancer,

CC particularly breast and ovarian cancer. The nucleic acid sequences,

CC proteins, agonists and antagonists may also be used in the diagnosis,

CC prevention and treatment of immune disorders e.g. Addison's disease,

CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,

CC

CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ Sequence 3008 BP; 792 A; 711 C; 737 G; 765 T; 3 other;

her;

Query Match	3.3%;	Score 200.6;	DB 21;	Length 3008;
Best Local Similarity	58.0%;	Pred. No. 8e-42;		
Matches 397;	Conservative	0;	Mismatches 279;	Indels 9;
				Gaps 2;

2.

QY	2668	TCGCGTGGTGGTGGGAATTCCTCCCAACCTGCAATCAGAGGACCTTCTGGGGCTGACATGG	2727
Db	1356	ctcgcttcctcggaaataacacgaataaaaaaccctccagaaagatctctctcggccaaaaacc	1655
QY	2728	GCGCTTCCCATCATGCTGGTGGTCCGCTACTGAGAGAACAGAGTTTGAACACCGATTTTACCTT	2787
Db	1656	gggtctccatataaagaatcgtcgaaggacccgctcgaaaaaacagttaaagctatctctccctt	1715
QY	2788	GGAATGAATTTGAGACAGAGAGCCATCTCTGTCATTTGATGAGCAT---GCTCACTCGGCC	2844
Db	1716	atgataaactcagacacagaagcctgctctcgcctatgatgatcatcatatcctatgctgaact	1775
QY	2845	ATGACGAAATCATGTTTGGTGGTTCCGGGTGTGGAGAGAACCTGGGAGCCGATTCGTGGCT	2904
Db	1776	ctcgaagcctcgcaaatcttggtatcgaagctcggcgggaaattcttcgaccggtctggcggtt	1835
QY	2905	TCCCTGGCCGCTTACACAGCATGGGAGACATCCCCATCAGTCTCGCTTACAACTCCACT	2964
Db	1836	accgggtctgctcgcatactctcgtgaccacagagatgaataagtcggaagctagctcgaagt	1895
QY	2965	ACTGCTGTGAGCTGTCCATGGTGTCTGACAGGTGCTGCTTCTTTACAAAGTATTAGCCT	3024
Db	1896	ggacgaatgaagtcgtccatcgtgctcacgcggggcagctttcatcaagaattttaact	1955
QY	3025	ACCTGTATTCTTATGTGATATGCCCCAGGCCATCCGGAGCATGTGGATGATATCATCAACT	3084
Db	1956	accgtctacctataaaaaatcgtccggggatataccaagaactcggctagaatcgtcatatgaact	2015
QY	3085	GTGAGGACATTGGCCATGAATCTTCTTGTGCCACATCACTGGAGAGCCGCCATCAAGG	3144
Db	2016	gtcgaagatatcgcacagaaactctccctcgtggtccaaagctcacgggaagaagctatcaagg	2075
QY	3145	TGACCTACAGCGTGTGACATTTCCGATGGCCAGAGATGCC-----TCAGGCCCTGTCTCATG	3198
Db	2076	taaccaccaagaagaataatcgaatgctcctcgaagtcgacacagcatagatggtctctcaact	2135
QY	3199	ATGACCTCCCACTTCCACGAGCGGCACACAGTGCACTCTTCTCGTGAAGGTGATACGGCT	3258
Db	2136	accaaacacacacatgctggaaggttcgaagtcgacatcaacaagttctctcaagctcgcga	2195
QY	3259	ACATGCCCCCTCTCTGTACACGCAATTTCAAGGTGTGATTTCTGTCTTCAAGACACGCCCTGC	3318
Db	2196	ccatgctctcttaagaatgctggaacacacgaactcgaaccctgctctgatacaagaatgaacttc	2255
QY	3319	CCCATGTACAAGACCACTGCTTTAA	3343
Db	2256	ctgagaagctcgaagaagcttccccaa	2280
RESULT 14			
AAZ08224			
ID	AAZ08224 standard; cDNA; 3175 BP.		
XX	AAZ08224:		
XX	25-JAN-2000 (first entry)		
XX	Human Ext-2 cDNA.		
XX	Human Ext-2 cDNA.		
XX	Human Ext-2 cDNA; Ext-2 protein; chromosome position 11p11-13;		
KW	hedgehog protein; hedgehog interacting protein; diffusion regulation;		

KM glycosyltransferase; GAG chain; protein core; proteoglycan;
KM hedgehog-mediated signal transduction; hedgehog specific GAG chain;
KM ext dependent pathway; drug discovery assay; growth factor; apoptosis;
KM cell proliferation; cell differentiation; cancer; ds.

OS Homo sapiens.

	Key	Location/Qualifiers
FH	CDS	335..2491
FT		

```
/note= "Human Ext-2 protein"
```

PD 07-OCT-1999.

PE 30-MAR-1999; 99WO-US06892.
XX

PR 30-MAR-1998; 98US-0079928.
XX

PA (HARD) HARVARD COLLEGE
XY

P1 Bellalache Y, The SI, Perrimon N;
XY

WPI; 1999-610841/52.
P-PSDB; 200300003

DR P-PSDB; AAY28883.

PT Identifying inhibitors of signal transduction of extracellular proteins, used for developing agents for treating, e.g. proliferative disorders or cancers or for cartilage -

Example 3; Page 99-103; 141pp; English.

Sequence 3175 BP; 780 A; 792 C; 817 G; 786 T; 0 other;

sequence 31/5 BP; 780 A; 792 C; 817 G; 786 T; 0 other;

Query Match

Query match	3.3%;	Score 200.6;	DB 20;	Length 3175
Best Local Similarity	58.0%;	Pred. NO. 8.2e-42;		
Matches 307; Conservativity	0.0	Mismatched 370		

2668	TCGTGGTGGTGGGAATTCCTCCCAAGCTGGCCATATAGAGGACCTTTGTGGCCCTGACATTGG	2727
1791	ttgtcgtcttgtaataatacaataataaaacccctccagaagattctcttgcgcaccaaatcc	1850
2728	GCCTTCCCATCATGGTGGTCCGTACTGAGAGAACAAGATTGAACAAOCGATTCTTCAACCT	2787
1851	gggttccattataaagttctgtgagactctgtaaacaagttaagtaaacggttcttcctt	1910
2788	GGAAATGAATTGAGACAGAGAGGCCATCTCTGTGCCATTGATGACGAT---GCTCACCTCTCGCC	2844
1911	atgaagaatactgacagacaagaactgtctctgtgccaatttgatagtatacatcatatctgaacct	1970
2845	ATGACGAATCATCTTTTGGGTTCCGGGTGTGGAGAGCAAGCTCGGAGCCGATGATGGGGCT	2904
1971	ctgacgagactgcaatttggtatagaggtctcgggagaaattcctgcaccggttggtyggtt	2030
2905	TCCCTGGCCGCTTACCACGCGATGGAGACATCCCCCATCATGCTCGGTCTTCAACTCCCACT	2964
2031	aaccgggttcgctctgcaactctctggagacatagatgaataagttgaaagttaagttcgaatt	2090

QY	2965	ACTCCTTGACCTCTCCATGATGGTGTGACAGGTGCTCCTTCTTTACAAATATATAGCT	3024
Db	2091	ggacgaatgaagatgctccatgagtgctcaatcggggagcctttatcaacaagatatttaact	2150
QY	3025	ACCTGTATTTCTTATGTGATGCCCCAGGCCATCCGGGACATGGTGTGATGATACATCAACT	3084
Db	2151	acctgtatacctatacaaatgcccgggagatatacaagaactcgggtagatgctataatgaact	2210
QY	3085	GTGAGGACATTTGCCATCAACTTCTTGTCTCCACATCACTGTGGAAAGCCCCCATCAAGG	3144
Db	2211	gtgaagatattgccaatgaactctctgtgtgccaacgctcacggaaagaacagatltcaagg	2270
QY	3145	TGACCTTACGGGTGACATTCCTCGATGGCCAGAGATGCC-----TCAGGCCCTGTCATCG	3198
Db	2271	taaccaccaagaagaatlttaactgaagtcctcgagtgacagcatalagatggctttcaactag	2330
QY	3199	ATGACTCCCACTTCCACGAGCGGCACACAGTGCATCAACTTCTTCTGTAAGGTGTAAGGCT	3258
Db	2331	accaacaacacatggtgtagaggttcaagtagtcatatacaaaagtltcttcagctcttgysga	2390
QY	3259	ACATGCCCTCTCTCTGTACAGCAGCACTTGAAGGTGATTTCTGTCTTTCAAGACACGCTGC	3318
Db	2391	ccatgacctctcaaggttggtgaaacacggactgacacctgtcctgtgacaaagaatgatcttic	2450
QY	3319	CCCATGACAGACCAAGTGTTCAA 3343	
Db	2451	ctgagaagctgaagagcttcccca 2475	

RESULT 15
NAV1037E

ID	standard; CDNA; 3003 BP.
AAV19375	

AC AAV19375

DT 20-AUG-1998 (first entry)

DE Hereditary multiple exostose associated EXT2 gene isoform encoding cDNA.

KW Hereditary multiple exostose; EXT2; chondrosarcoma; human; isoform
KW treatment; ss.

OS Homo sapiens.

FH	Key	Location/Qualifiers
100		100
101		101
102		102
103		103
104		104
105		105
106		106
107		107
108		108
109		109
110		110
111		111
112		112
113		113
114		114
115		115
116		116
117		117
118		118
119		119
120		120
121		121
122		122
123		123
124		124
125		125
126		126
127		127
128		128
129		129
130		130
131		131
132		132
133		133
134		134
135		135
136		136
137		137
138		138
139		139
140		140
141		141
142		142
143		143
144		144
145		145
146		146
147		147
148		148
149		149
150		150
151		151
152		152
153		153
154		154
155		155
156		156
157		157
158		158
159		159
160		160
161		161
162		162
163		163
164		164
165		165
166		166
167		167
168		168
169		169
170		170
171		171
172		172
173		173
174		174
175		175
176		176
177		177
178		178
179		179
180		180
181		181
182		182
183		183
184		184
185		185
186		186
187		187
188		188
189		189
190		190
191		191
192		192
193		193
194		194
195		195
196		196
197		197
198		198
199		199
200		200
201		201
202		202
203		203
204		204
205		205
206		206
207		207
208		208
209		209
210		210
211		211
212		212
213		213
214		214
215		215
216		216
217		217
218		218
219		219
220		220
221		221
222		222
223		223
224		224
225		225
226		226
227		227
228		228
229		229
230		230
231		231
232		232
233		233
234		234

/*tag= a

PN EP837127-A2

PD 22-APR-1998

PF 26-AUG-1997; 97EP-0306503.

PR 21-OCT-1996; 96CN-0121928.

PA (UYHU-) UNIV HUNAN MEDICAL.

PI Deng HX, Fan CH, Xia J,

DR WPI; 1998-219110/20

DR P-PSDB; AAW44851.

PT Cloned human EXT2 gene - associated with hereditary multiple

PS Claim 4; Pages 28-31; 31pp; English.

CC This cDNA encodes an isoform of the EXT2 gene associated with hereditary
CC multiple exostoses and Chondrosarcoma. The polynucleotide is an isoform
CC of the EXT2 gene described in Nature Genet., 14, 25, 1996. The
CC polynucleotide can be used in the detection and treatment of EXT2-related

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 4, 2002, 14:57:01 ; Search time 26.04 Seconds
(without alignments)
3391.171 Million cell updates/sec

Title: US-09-809-920-4

Perfect score: 4873
Sequence: 1 MTGYTLNRNGAGNGGQTCM.....DSVLEKTRLPMDKTKCFKFI 919

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4873	100.0	919	2 JC5934	exostose-related p
2	1252.5	25.7	814	2 T23200	hypothetical prote
3	367	7.5	330	2 JC5935	exostose-related p
4	360.5	7.4	334	2 T49195	hypothetical prote
5	306.5	6.3	764	2 T48446	hypothetical prote
6	288	5.9	329	2 D96834	hypothetical prote
7	232.5	4.8	444	2 T20803	hypothetical prote
8	123.5	2.5	425	2 T05631	hypothetical prote
9	121.5	2.5	511	2 T51544	hypothetical prote
10	120.5	2.5	680	2 AG2419	hypothetical prote
11	120.5	2.5	1202	2 S55553	LAR-interacting pr
12	120	2.5	450	2 F84764	hypothetical prote
13	120	2.5	893	2 F72253	hypothetical prote
14	119.5	2.5	824	2 I50618	c-fps proto oncoge
15	118.5	2.4	622	2 S45129	VPS27 protein - ye
16	117	2.4	440	2 C96697	hypothetical prote
17	117	2.4	440	2 T02165	hypothetical prote
18	117	2.4	1331	2 A72647	probable surface 1
19	115.5	2.4	811	2 JC7619	hypoxia-inducible
20	114.5	2.3	1024	2 G71434	probable limonene
21	114	2.3	670	2 T38446	microtubule-associ
22	113.5	2.3	385	2 H83930	lipase (esterase)
23	113.5	2.3	359	2 S74668	hypothetical prote
24	113	2.3	676	2 F69394	H+-transporting AT
25	113	2.3	1637	2 AE2109	two-component hybr
26	113	2.3	2633	1 S28261	centromere protein
27	112.5	2.3	4464	2 D87755	protein T21E12.4
28	111.5	2.3	2484	2 T26216	hypothetical prote
29	111.5	2.3	2607	2 T26215	hypothetical prote

30	111.5	2.3	3075	2 S14458	lamin alpha-1 ch
31	110.5	2.3	382	1 I40209	cytochrome p450 B7
32	110.5	2.3	1940	2 A29320	myosin heavy chain
33	110.5	2.3	5138	2 B96695	hypothetical prote
34	110	2.3	1207	2 T13827	kinesin-73 - fruit
35	109.5	2.2	1207	2 B88789	protein ZK1251.9
36	109.5	2.2	1596	2 A41216	guanine nucleotide
37	109	2.2	475	2 T47480	hypothetical prote
38	109	2.2	765	2 E96558	hypothetical prote
39	109	2.2	978	2 A70387	conserved hypothet
40	109	2.2	4967	2 S72269	ryanodine receptor
41	108.5	2.2	346	2 T48520	hypothetical prote
42	108.5	2.2	779	2 AG1978	hypothetical prote
43	108.5	2.2	1212	2 T00332	hypothetical prote
44	108	2.2	720	2 D98141	bme12 protein (Af0
45	108	2.2	720	2 AF3146	polysaccharide blo

ALIGNMENTS

Query Match	Best Local Similarity	100.0%	Score 4873	DB 2	Length 919	Matches 919	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY 1	MTGYTLNRNGAGNGGQTCMLRMSNRIRLTWLSPTLVILVFPPLIAHYVLTTLDEADEA 60									
Db 1	MTGYTLNRNGAGNGGQTCMLRMSNRIRLTWLSPTLVILVFPPLIAHYVLTTLDEADEA 60									
QY 61	GKRIFGRVGNELCEVKAHVLDLCIRISVSEELQLEAKROELNSETAKLNKTEACKS 120									
Db 61	GKRIFGRVGNELCEVKAHVLDLCIRISVSEELQLEAKROELNSETAKLNKTEACKS 120									
QY 121	IENKODLQKNTVISOETHSYKELMAONOPKLSLPIRLLPKODAGLPPEKATRGCRRLH 180									
Db 121	IENKODLQKNTVISOETHSYKELMAONOPKLSLPIRLLPKODAGLPPEKATRGCRRLH 180									
QY 181	NCFDYRSCPLTSGPVPVYVSDQFVFGSYLDPVKAQFOATARAVVVTENADIACLYVI 240									
Db 181	NCFDYRSCPLTSGPVPVYVSDQFVFGSYLDPVKAQFOATARAVVVTENADIACLYVI 240									
QY 241	LVGMQEPVYLPALEKOLYSLPHMRTDGHNVITLMSKSDTQNLVNVSTGRANVAQ 300									
Db 241	LVGMQEPVYLPALEKOLYSLPHMRTDGHNVITLMSKSDTQNLVNVSTGRANVAQ 300									
QY 301	STFTVQYRPGFDLVVSLVHAMSEPMFETPPQVPRKRYLTFQGEKTESLRSSLOEA 360									
Db 301	STFTVQYRPGFDLVVSLVHAMSEPMFETPPQVPRKRYLTFQGEKTESLRSSLOEA 360									
QY 361	RSFEEMEGDPADYDRRIATLKAQVDSKLDQVLEFTCKNOPKPSLPTEMALCGERED 420									
Db 361	RSFEEMEGDPADYDRRIATLKAQVDSKLDQVLEFTCKNOPKPSLPTEMALCGERED 420									
QY 421	RLELKSTFALLITTPGDPRLVYSSGCATRLFEALVGAAPVVLGEVOLPYQDMLOWNE 480									
Db 421	RLELKSTFALLITTPGDPRLVYSSGCATRLFEALVGAAPVVLGEVOLPYQDMLOWNE 480									

Db 421 RLEKLKSTFALLITTPGPRIVISSGCATRLFEALEVGAVPVVLCGOVOLPYQMLQWNE 480
Qy 481 AALVVPKPRVTEVHFLSLSDSDLLAMRROGRFLMETEYFSTADSIPTVLMIRTRIQI 540
Db 481 AALVVPKPRVTEVHFLSLSDSDLLAMRROGRFLMETEYFSTADSIPTVLMIRTRIQI 540
Qy 541 PAAPRREAAAEIPRRSKAGTDPNMADNGDLDGPEVTEPPYAPRILNFTLTVD 600
Db 541 PAAPRREAAAEIPRRSKAGTDPNMADNGDLDGPEVTEPPYAPRILNFTLTVD 600
Qy 601 YRSMNACGPFHLPFHTFDDVLPSEAKFLSGTFRPFGAGSGGKEFOALGCVNR 660
Db 601 YRSMNACGPFHLPFHTFDDVLPSEAKFLSGTFRPFGAGSGGKEFOALGCVNR 660
Qy 661 EGFVYVMTYREEVYVMSLERLNGLPYLNKVVVYVNSPKLPSEDLMPDVGVPIMVVRT 720
Db 661 EGFVYVMTYREEVYVMSLERLNGLPYLNKVVVYVNSPKLPSEDLMPDVGVPIMVVRT 720
Qy 721 EKNSLNNEFLPMNEIETFAILSIDDAHLRDEIMFGFRVWRREADRIVGPGGRHAMD 780
Db 721 EKNSLNNEFLPMNEIETFAILSIDDAHLRDEIMFGFRVWRREADRIVGPGGRHAMD 780
Qy 781 PHQSMVYNSYSCSELNMTLGAFFHKYAYLYSVMPQAIRDMVDEYINCEDIAMNPLY 840
Db 781 PHQSMVYNSYSCSELNMTLGAFFHKYAYLYSVMPQAIRDMVDEYINCEDIAMNPLY 840
Qy 841 SHTRKRPPIKVTSMWTFRCPCPOLSHDSDHFERHKCINFVKVGYMPLLYTQFRVD 900
Db 841 SHTRKRPPIKVTSMWTFRCPCPOLSHDSDHFERHKCINFVKVGYMPLLYTQFRVD 900
Qy 901 SVLFKTRLPKDKCKEFT 919
Db 901 SVLFKTRLPKDKCKEFT 919

RESULT 2

T23200
hypothetical protein K01G5.6 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Sep-2000
C/Accession: T23200
R/Fall, M.
submitted to the EMBL Data Library, March 1997
A/Reference number: Z19706
A/Accession: T23200
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-814 <NLI>
A/Cross-references: EMBL:Z92803; PIDN:CAB07245.1; GSPDB:GN00021; CESP:K01G5.6
A/Experimental source: clone K01G5
A/Genetics:
A/Map position: 3
A/Intons: 7/1: 52/2: 112/1: 156/2: 216/3: 702/3: 753/2
C/Superfamily: Caenorhabditis elegans hypothetical protein K01G5.6

Query Match 25.7%; Score 1252.5; DB 2; Length 814;
Best Local Similarity 33.4%; Pred. No. 5.5e-82;
Matches 313; Conservative 147; Mismatches 312; Indels 165; Gaps 25;

Qy 21 LRMSNRRLRLTWLSTFLVILVFPPL--IAHYLLT-----TLDEADGAKRI 64
Db 5 LNGSSRSFVPSLRVSALIFVITIIYIYNVFSFSEPSWITQDALKONIENTLDYDASC 64
Qy 65 FGPRVGNELCEVKKHVLDCRRESVSE-ELQLLEAKROELNLTAKLNKLEACKKSTEN 123
Db 65 SGVSTIGILRQKRIILASVRLTESOVKIEIRTVQEEQLRLPKOKLELSALEGETEA 124
Qy 124 AKODLLQKNNVISOEHSYKELMAQONPKLSLPIRLPEKDDAGLPPEKATGCRNLNCF 183
Db 125 AQORL-----ELRKTQNVKVFLLPFS--PLQIRLELDPQSQISPNQOLDIT 168

Qy 164 DYSRCPITSGFPVYVYDSDQVEGSLDPLVKQAFQAFARANVYVENADIACTLYILVG 243
Db 169 DYSRCSISSFMPYV----- 183
Qy 244 EMQGEVVLRLPALEQOYLSLPHMRTDGNHVINLSRKSPTQNLNLYSTGRAMAOSTF 303
Db 164 -----DITSGSEKEMLV-----FOEYVPLNLETPDKACIHIHNSNGIAS--PNTTF 230
Qy 304 YTVQRPFGDVLVSPVLHAMSEPNMEIPQVPRKYLTFQGEKISLSSLOEARSF 363
Db 221 NSILEFVG-----SPLIN-----FQSSIHVQSKI--RSF 259
Qy 364 EEMEGDPPADYD---DRI-IATLKAVQDSKLDVLYEFTCKN-----QPKPS-LPT 410
Db 260 -----DFPVVYNHIAVAKVDLPLLPQRENLSLIDNTELENSAFSSLSAEPSSRPPI 313
Qy 411 EMALGGER-----EDLELKLKSTFALLITTPGPRIVISSGCATRLFEALEVGAVPVVIGE 466
Db 314 VIYKCSQENCSLERRROLIGSTFCFLP-----SEMFQDFLSLQJGCIPIILSN 365
Qy 467 QVOLPYQDMLQWNEALVVPKPRVTEVHFLSLSDSDLLAMRROGRFLMETEYFSTADI 526
Db 366 SQLPFPQDLIDMRATYRLPLARLPEAHFIQSPESIDIIEMRRYGRLEFYETYLADRI 425
Qy 527 ENTVYAMTRTRIQIPAAPIREBAAEIPRRSGKA-AGTDPNMADNGDLD--LGEVTEPP 583
Db 426 ARSLAALRLYKLIQIPTRVRRNOALPLENSSFTAPKGSVVVQANFDEYILGPLSR-- 483
Qy 584 YASPRYLNFTLVVDFEYRSMNACGPFHLPFHPFDVLPSEAKFL-GSGTGFPRIGG 642
Db 484 VESTSYANFTFEOQLYSTDFNNITMSPHYTEFLVNAELPTEAEFFEDTIGFRPI-- 540
Qy 643 AGSGKEFOALGCVNVPREOFTVYMLTYEREVLNMSLERLNGLPYLNKVVVYVNSPKLP 702
Db 541 EPGSGAEFSKALGNGRQREOFTVYMLTYERDAVILGALERHQLPYLNKIIIVYVNNVARD 600
Qy 703 SEDLMPDYGVPINVYVPRKSNLNRLPANEIETFAILSIDDAHLRDEIMFGFRVWR 762
Db 601 PPD--TWPSLHLPVEFIRVAENLNRRFVPMRIETEAVALSIDDDIDLMQOEIIIAFRWR 659
Qy 763 EARDRIYGFPGRYHAMDIPHQSMVYNSYSCSELNMTLGAFFHKYAYLYSVMPQAIR 822
Db 660 ENRDRIYGFPRHHA--RYGDSMFYNSNHTCOMSLTGAFFHKYAYLYSVMPQAIR 717
Qy 823 DMVDEYINCEDIAMNFTLSTRKRPPIKVTSMWTFRCPCPOLSHDSDHFERHKCINF 882
Db 718 EHVNSIKNCEDIAMNYLVSLTRKRPPIKVTSMWTLKCTCIESLYKGETHEKHECRL 777
Qy 883 FVKVYGYMPLLYTQFRVDYVLFKTRLPKDKCKEFT 919
Db 778 FTKIYGYNPLKFSQFRADSLFKTRLPQNHQCKEYV 814

RESULT 3

JC5935
exostose-related protein 2 - human

C/Species: Homo sapiens (man)
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C/Accession: JC5935
R/Saito, T.; Seki, N.; Yamauchi, M.; Tsuji, S.; Hayashi, A.; Kozuma, S.; Hori, T.
Biochem. Biophys. Res. Commun. 249, 61-66, 1998
A/Title: Structure, chromosomal location, and expression profile of EXTR1 and EXTR2,
A/Reference number: JC5934; MUID:98139867
A/Accession: JC5935
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-330 <SAI>
A/Cross-references: DDBJ:AB009284; NID:92723392; PIDN:BAA24081.1; PID:92723393

Query Match 7.5%; Score 367; DB 2; Length 330;
Best Local Similarity 30.8%; Pred. No. 1e-18;
Matches 86; Conservative 49; Mismatches 98; Indels 46; Gaps 7;

```

OY 656 GAVPEOEFVWVLTEREVEVLNMSLERLNGCLPYLNKVVVVVMSKPLSDELDPDIG--- 712
D 59 GKSTMSDFLLIMOTNRNRDLDLKLNLHYQAVPRLKHVYVWNNGEAKADELMSLGP 118
OY 713 VPIVWVTEKSNLNRFLPEWNEIETFEALISIDD AHLRDEIMFGFRVWRREARDIVGFP 772
D 119 IPIVKOOTANMRNRRLQVFPPELTNAVYLMVDDDTLISPDLEAFASVMOOPFDQIVGV 178
OY 773 GRVHAMDPHOSMLVN-----SNVSCELSVLTGAAFPKHYAYLYSVVMOAIR 822
D 179 PRKH---VSTSGCISYSGFEMQACSGNGDOYSWYLIGASFNSKYLELQF-QRAVH 234
OY 823 DWVDEYINCEDIAMNFLVS-HITRKPPIVY-----TSRWTFRCPCGPQA 865
D 235 ALIDTQNCDDIAMFEIIRKHIKGTSGIETVFKPVYMNMLEKETNSGSMV----- 284
OY 866 LSHDDSHFERHKCNFVFKVGYGMPLYTQFVDSVLF 904
D 285 --HRAEHALORSYCINKLVNIYDSMPLRYSNIMISQFGF 321

```

RESULT 4
 T49195
 hypothetical protein F27K19.10 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
 C:Accession: T49195
 R:Bernes, V.; Wurmbach, E.; Dzonek, H.; Ansoerge, W.; Mewes, H.W.; Rudd, S.; Lemcke, K.,
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: Z25014
 A:Accession: T49195
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-334 <BEN>
 A:Cross-references: EMBL:AL163832; GSPDB:GN00061; ATSP:F27K19.10
 A:Experimental source: cultivar Columbia; BAC clone F27K19
 C:Genetics:
 A:Gene: ATSP:F27K19.10
 A:Map position: 3
 A:Introns: 71/2, 203/3, 258/2, 282/1

```

Query Match          7.4%: Score 360; DB 2: Length 334;
Best Local Similarity 33.5%: Pred. No. 3.3e-18;
Matches 93: Conservative 41; Mismatches 88; Indels 56; Gaps 7.

660 REQFVVMVLTYYREYVLMNLSLERLNGDLPYLNKVVVVYVMSPKLPSEDL----- 706
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
71 RKGYTLMLMTMKRRYDLKKSVSHVASCRLSDIHIVMSSPNPSESLEKYLHNVLKKKTR 130
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 707 -----LMPDGVPIVMVYTERKNSLNPNRELPMNEITEALSIDDAAHLRHDEIMGFERY 760
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 131 DGEYELAFDI-----NKEDSLNNRKEKELDKTDVAVFISDDIITPCGHTVDAFNV 182
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 761 WREARDRIYGFEGRYHANDIPHOSMLYSN-----YSCELSMVLTGAFF 805
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 183 WESAPDITWGVFPRVH-----WPEKSNDAKANYTYSGMMSVMMSCGYSLVLSAAAF 234
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 806 HKYVYLYSYVPAQIRDMVDEYINCEDIAMNLFVLSHITRKPIKITSR-WTFRCGCPQ 864
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 235 HKKYLSTLNSPASIREFTTKNRNCEDIAMSFLLANATNAPAIWKGKIYEIGSG--- 291
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 865 ALSHDSHFHEHKKCINFPKYVGYGMPLLLYTQFR-VDS 901
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 292 -ISSIGGHTKEKRTNHCYNRFAVEFGKMPILYYTSMKAVS 328
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

RESULT 5
T48446
hypothetical protein T32M21.100 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T48446
R:Beran, M.; Terry, N.; Ardiles, W.; Buyschaert, C.; Dasseville, R.; De Clerck, R.; R. ...
ees. H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z2487
A:Accession: T48446
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-764 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 105/3; 146/3; 202/3
A:Note: T32M21.100

Query Match	Similarity	6.3%	Score 306.5	DB 2	Length 764
Best Local	Similarity	33.5%	Pred. No. 9e-14		
Matches	91	Conservative	49	Mismatches	95
				Indels	37
				Gaps	11

QY	635	GFRPIGGAGG-SGKEFOALGQNVPRDFOVTYMLTYE-REEVYLNLSLERLNGLPYLNKV	692
DB	494	GVRYLYYGSGAEPYPRFKHL-----SQFTLATMYDARLNLMKYVRYKRSCEVSKEI	547
QY	633	VYVWNSPKRPSEDLMPDLPVPIVYVTRKSNLNNFLEPWNLETFEALISIDDAHLRHD	752
DB	548	VIVWV-KGPPDLSBLSAIVPVRIYVQKQNSLNNFELDLIKTRAVLELDDDDIMPCD	605
QY	753	ELMGEFRVWRARDRIVGPPGRYHAMADIPHQSMLYNSNTSC-----LSNVLTGAAFFHKY	808
DB	606	DEKEFRVWRREPRERLYGEPYPRF-----VDQMTYSAEKFAARSHKGYNMLTGAAFMVDR	660
QY	809	YAY-LYSYVMPAIDMVDYDECYDIAMNPLVSHSTR-----KPIKY--TSRMTF	857
DB	661	FAFDMYQSDKALIGVAVDEQFNCEDLILNPLIYANASGCAVEYVRPSLVITIDISKFS-	719
QY	858	KCPGCPQALIS-HDSDHFHRHKCINFEVYVYG	888
DB	720	-----GVALISGNTNQHYYRKRKSKLRRFSDLYG	746

RESULT 6
D96834
hypothetical protein F516.4 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96834
R:Rheologis, A.; Ecker, J.R.; Palm, C.J.; Federpiet, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Hutzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Maiti, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallia
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719
A:Accession: D96834
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-329 <STO>
A:Cross-references: GB:AE005173; NID:g6751704; PID:AAF27686.1; GSPDB:GN00141
C:Genetics:
A:Gene: F516.4
A:Map position: 1

Query Match	5.9%;	Score 288;	DB 2;	Length 329;
Best Local Similarity	27.7%;	Pred. NO. 5e-13;		
Matches	79;	Conservative	52;	Mismatches 106;
			Indels	48;
			Gaps	8;

Db 538 MADGHTD-----SYSSSAVLRRRPQKRLAALRDEPSKVOTLNEQDMERAQASVLANVAQ 592
Qy 454 ALEVAVPVVIGE---QVQLPYQDMLQWNEALVVPKPRVTEVHFLRLSLSDSLLAMR 510
Db 593 AEDSDA-DVSDGEDDRDRLTLSSVDLLSPSGNA-----DAHTLAMLQe-QDAINK 641
Qy 511 OGRFMEYFEST---ADSIENFV-----LAMIRTRIQIPAPPIREEMA- 551
Db 642 EIRLIOEKENTEQAAEIESHVSGSLDNLGRFRSMSSIPYPASSSLASSSPSGRST 701
Qy 552 --EIPH-----RSGKAAGTDPNMAD-NGDLDLGPVETEPPEYAPRYLRNFTL----- 595
Db 702 PRIPHSPAREVDRGVMTLLPSPSREVRDOKTITKETSPP-SSPRALRLDRLLKGAH 760
Qy 596 ----TVDFYRSNMCACGP-----FHLPHFPDPVLPSEAKFLGSGTGR 637
Db 761 TVSHEDIDIRNSGSDGPVSNPSSNSQDSLHKAP-----KKKQIKSSIG-R 809
Qy 638 PIG---GGAGSGKEF---QAA-----LGGVNPREFQTVVMTYEREDV 675
Db 810 LFGKKERKRGPGOTGREALGQAGVSETDNSQDALLGSLKLGQAEKRN-----KLOKKHEL 864
Qy 676 LMSLERLNGLPYL---NKVYVYVNSPKLPSEDLMPDIGVPIVYVTEKNSLNRFPLM 732
Db 865 LEEA--RRQGLPFAQMDGPVYVWLE-----LM-VGMPAMYVACRAN----- 904
Qy 733 NEIEFALSTIDDAHLRHD 752
Db 905 --VKSALMSALDSTEIORE 922

RESULT 12

F84764

hypothetical protein Atg35100 [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cross)

C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C/Accession: F84764

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shee, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: F84764

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1460 <STO>

A:Cross-references: GB:AE002093; NID:g3668093; PIDN:AAC61825.1; GSPDB:GN00139

C:Genetics:

A:Gene: Atg35100

;Map position: 2

Query Match 2.5%; Score 120; DB 2; Length 460;
Best Local Similarity 19.6%; Pred. No. 1.1;

Matches 105; Conservative 79; Mismatches 162; Indels 190; Gaps 26;

Qy 146 MAONQPKLS-LPRLRLPEKODAGLPPEKATRGRLHNCFDYSRCPLTSGFPVYVD-SDQ 203
Db 29 VSRSLPSSSDLPRLIREDD-----DEGRAPIDQPRVAVMYNLPKR 70
Qy 204 FVFGSLDPLVKAQFATARN----- 225
Db 71 FYTG-----LIEQ-HSIAGGIKKPVGDVTLTKYPGHMHMYLFSDLNQPEVDRSGS 123
Qy 226 --VYVENADIACTVYLIVGEMQEPV-VLRAPEL-----EKOLYLPHM-----RT 268
Db 124 PLYRVSDDPADLFFYPVSSLSLIYNAGRPRVAGSGYSDKQECILVLEQDEMMWRN 183
Qy 269 DGHNVHITLMSRKSQDNTLQNLVNV-STGRAMVAQSTFYTYQYRPPDL--VSPPLVHAM 324
Db 184 AGHDHYI-----PAGDPNMLYRILDRVKNKNAVLLVSDFGRLRPQGSFVADVVIPIYSHRYA 238

Qy 325 EFNEMEIPQVPEV-RKYLTFEÖGKIESLRSLQEAANSFEEEMEGDPPADDRITATL 383
Db 229 IEN-----GEIGVEDRNTLLFFMGNRKR-----DGKVRDLDFVL----- 275
Qy 384 KAVQDSKLDQVLYEFTCKNQKRPSPLEPMALGGERDELLELK-LSTALITP-GDRL 441
Db 276 -----EKEDDVYIKHGTS-----RENNRAATKGHTSKFCINPAGD--- 312
Qy 442 VISSGCATRLFEALVGAVPVVLGBOVLPYQDMLQWNEALVVPKPRVTEVHFL-----LR 498
Db 313 ---FTSACRLDPSIYSLCVPLIVSDIILPEVDYIDYKFSIFVANAALDPGLVOMLR 369
Qy 499 SLSDSDLLAMRRÖGRLMEYTFSTADSIFNTVLAMIRTRIQIPAPPIREEMAAETPRSG 558
Db 370 KIKTKILEYÖREMKSKD-----FTKLMSM-----HK-- 398
Qy 559 KAAGDPMMAONGDLDLGPVETEPPEYAPRYL--RNFLLTYTDFYRSNMCAPGPH 612
Db 399 -----PNRSYHFLNLNLFVERPFGMLRSLRRREFTEYTSRY-----SPLPMH 443

RESULT 13

F72253

hypothetical protein TM1450 - Thermotoga maritima (strain MSB8)

C/Species: Thermotoga maritima

C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C/Accession: F72253

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316

A:Accession: F72253

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1893 <AAN>

A:Cross-references: GB:AE001796; GB:AE000512; NID:g4982004; PIDN:AAD36518.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM1450

Query Match 2.5%; Score 120; DB 2; Length 893;
Best Local Similarity 19.3%; Pred. No. 3.3;

Matches 107; Conservative 84; Mismatches 194; Indels 170; Gaps 24;

Qy 247 EPVYLRAELEKÖLYSLPHMRTDGHNVITNLSKSDTQMLTN--VSTGRAMVAQST-- 302
Db 30 EDVVLSPKIKRIEIL--WRLNGENLKIATTLKALTEKVFSPDFLRNSLMIASTKL 87
Qy 303 -----FYVQ-----YRGG--FDLVVSPLVHAMSEPNMEIP 332
Db 88 TLSPEKIVEMGYELVTFVQVANGFAIRGLIDIIYSPDNPPVRIELFGDEIEIRFFKV 147
Qy 333 PQVVPVKKRY-----LTFÖGKE-----IESLRSLQEAANSFEEEMEG 369
Db 148 TQ-----RSFGVMDKTLILPVDYGGSTLIDFLKTARFICEDLQVIVDEYRKFRKEMRD 203
Qy 370 DPPADYDDRIATLKAQVDSKLDQVLYEFTCKNQKRPSPLEPMALGGERDELLELKLT 429
Db 204 -----LLKERYNDFDERVVEVLKAVKESAPLSTRV--EKKESLPILDVDE 249
Qy 430 FALITPGDPRIVYISSGCATRLFEALVGAVPVVLGBOVLPYQDMLQWNEALVVPKPR 489
Db 250 IE-----BGLVNHREHGIAIEG--IVRLKGVILGERYL-----KLKEDALLVPIEK 297
Qy 490 VTEVHFLRLSLSDSDLLAMRRÖGRLM-ETVYSTADSIENFNTVLAMIRTRIQIPAPPIRE 548
Db 298 IDVNH--KYIGDPSQVKLDRMNRGKWKOTLKKVREDIEKKIKELVELYK-----RÖE 348
Qy 549 AAALPIHRSKAGATDPNMADNGDLDLGPVETEPPEYAPRYLRNFTLYTDFYRSNMCAP 608

```

Db 349 A-----QGLSLRGPDELEK-----AESFYIETPOQOSIEEVLSD----- 386
Qy 609 GPHLPHTPPDPVLPSEANFLGSGTGRPIGAGSGSGKEFOALGANNPREQFTVVM 668
Db 387 -----LASEKPMDRLLCGDA-----GVKTEVALLRAAFRAVAVSGQVAVLVP 428
Qy 669 T-----YE-----REVLNMS-----LERLNLPLPNKVVVWNSPKL 701
Db 429 TTVLAKOHYENKFERMEREPGVKVELLDSSRTAREKKEITELK-----KGEIDIIITHTSL 484
Qy 702 PSEDLMPDGVPI 716
Db 485 LNERIEFSDGLVIT 499

RESULT 14
150618
C:fps proto oncogene - chicken
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 18-Jun-1999
C:Accession: 150618
R: Huang, C.C.; Hammond, C.; Bishop, J.M.
J. Mol. Biol. 181, 175-186, 1985
A:Title: Nucleotide sequence and topography of chicken c-fps. Genesis of a retroviral on
A:Reference number: 150618; MUID:85100839
A:Accession: 150618
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-824 <HUG>
A:Cross-references: EMBL:X02266; NID:g63203; PIDD:CA26155.1; PID:g871043
C:Genetics:
A:Gene: c-fps
A:Initons: 75/3; 133/3; 166/1; 227/2; 273/2; 313/2; 354/2; 416/3; 442/3; 512/3; 553/3; 5
C:Superfamily: protein-tyrosine kinase fps; protein kinase homology; SH2 homology
C:Keywords: ATP
F:462-547/Domain: SH2 homology <SH2>
F:561-823/Domain: protein kinase homology <kin>
F:569-577/Region: protein kinase ATP-binding motif

Query Match 2.5%; Score 119.5; DB 2; Length 824;
Best Local Similarity 18.8%; Pred. No. 3.2;
Matches 162; Conservative 114; Mismatches 291; Indels 293; Gaps 41;

Qy 79 VLDCIRTSVSEELDL-----BAKROELNSETAKNLKTEACKSTENAKODLLQ 130
Db 110 IRDQOLRAKFAFSEQMOOLSEYARTTOQEMEKTKAQRSLVRDSTQAKRQKQASKDER 169
131 LKNVISTEHSYKELMANONPKLSPIRLPEKDDAGLPPKATRGCRHLNCFYSKCP 190
Db 170 EKAREKTVRSIMKLYALHNDVYL-----AVRAAALHHHHHYORALP 210
Qy 191 TSGPVVYVYDSDF-----VFGSY--LDPLVKAQFQATARNAVVTENADIACLYLV 242
Db 211 TLHSLSKLSLOEMVLYLKEILGECSISSLVQEDVLAIHQEVNAHVEIDPATYSSFVQ 270
Qy 243 -----GEMQEPVLLPAEL--EKQDYSLEPMWTDGHNHVIINLSKRS 283
Db 271 CHRDESVPAVTPDESILLETESLEPGEQLNLETIESVQHSJLSTIEELLASREAVSS 330
Qy 284 TONLLVNVSTGRNAVASTFTTVQYRGRFDLVSPVYHNSSEPMELPROVPVKRYTLF 343
Db 331 KEQRYVMEIQ-----VELR-GEELALSP----- 351
Qy 344 TFOGKIESTL--RSSLOEARSFEEMEGDPPADYDRHILITLKAQV-----SKLDQVLEF 398
Db 352 ---GERVHLLGKRGLODA---QOOLGVLCAQ-----AKLOQRMLANKLAELISE- 398
Qy 399 TCKNQPRPSLPTL---MALGGERDRLELLKLSFTALITPGDRVLISSGCATRLFEAL 455
Db 399 ---EPPALPLQEDROSVCSTDERSGVTALETIKNHI-----SCIFSPRSL- 443
Qy 456 EVGAVPVVLEGOVLPYODMLQNHQALVLPKPRVTEVYHLLRSLSDLLAMRQGR-- 513

```

```

Db 444 ---PPVPLIPEVOKPLCQO-AMTHGAI-----PR-SEVQELK--CGSGDLVRESQKOE 492
Qy 514 -----FLWE-----TYFSTADSI-----FNTVLAIRTRIQIPAPIREEAAAEIPIRS 557
Db 493 YVLSVMDGQPRHFIIOAONLVRLEBDGFTPTPLDLHLLQ-----SQQPIRKS 543
Qy 558 G---KAAGTDPNMANGDIDLQ-----PV-----ETEPYAS 586
Db 544 GIYVLTAVLKDKNVLMNEEDVLGRIGRNGFEVSGRLRADNTPVAVKSCRETLPELK 603
Qy 587 PLYLRNFTLVTFDFFRSMNCAPGFHLFPHPPDPVLPSEAKFLGSGTGRPI----- 639
Db 604 AKFLQEARL-----LKQYN-----H-----PNIVRLGVCTQKQPIYVMELV 641
Qy 640 -GGGAGSGSGKEFOALGANNPREQFTVVMVLTREEVNMSLERLNGLPYLNKVVVWNS 698
Db 642 QCG-----DFLSFLRSEGR-----LKKKELIKMENNAAAGMEYLE----- 677
Qy 699 PKLPSEDLMPDGVPIVYVVRTEKNSLNRPVWNEIETAILSIDDAHLRHDEIMEGF 758
Db 678 ---SKHCIRHDLAARNCLV-TEKNT-----LKISD-----F 704
Qy 759 RVWREARDRIYVGFGRYH-----AMDIPHOSMLVNSYSCLSWLTGAFFHKYY--AYL 812
Db 705 GMSRQEDGVYASTGKAKQIPVKWTAPBA--LNYGRVSSSDVWSFGILLMEAFSLGAVP 762
Qy 813 YSYVMPQAIQDWDVETYNCE 832
Db 763 YANLSNQOTREAIQGVRL 782

RESULT 15
545129
VPS27 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein N2038; protein YNR006w
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence_revision 02-Aug-1994 #text_change 23-Mar-2001
C:Accession: S45129; S48343; A57274; S59736; S63332
R:Verhasselt, P.; Aert, R.; Voelt, M.; Volckaert, G.
Submitted to the EMBL Data Library, January 1994
A:Description: Twelve open reading frames revealed on the 23.6 kbp segment flanking t
A:Reference number: S45119
A:Accession: S45129
A:Molecule type: DNA
A:Residues: 1-622 <VER>
A:Cross-references: EMBL:X77395; NID:g496717; PIDD:CA545474.1; PID:g496723
R:Verhasselt, P.; Aert, R.; Voelt, M.; Volckaert, G.
Yeast 10, 1355-1361, 1994
A:Title: Twelve open reading frames revealed in the 23.6 kb segment flanking the cent
A:Reference number: S48338; MUID:95208356
A:Accession: S48343
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-622 <VE2>
A:Cross-references: EMBL:X77395; NID:g496717; PIDD:CA545474.1; PID:g496723
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1994
R: Piper, R.C.; Cooper, A.A.; Yang, H.; Stevens, T.H.
J. Cell Biol. 131, 603-617, 1995
A:Title: VPS27 controls vacuolar and endocytic traffic through a prevacuolar compartm
A:Reference number: A57274; MUID:96042307
A:Accession: A57274
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-320, 'NV', 323-622 <PIP>
A:Cross-references: GB:U24218; NID:g785066; PIDD:AAA96002.1; PID:g785067
R: Piper, R.C.; Cooper, A.A.; Yang, H.; Stevens, T.H.
Submitted to the EMBL Data Library, April 1995
A:Reference number: S59736
A:Accession: S59736
A:Molecule type: DNA
A:Residues: 1-320, 'NV', 323-622 <PIW>
A:Cross-references: EMBL:U24218; NID:g785066; PID:g785067

```

R:Aertl, R.; Verhasselt, P.; Voelt, M.; Volckaert, G.
 submitted to the Protein Sequence database, April 1996
 A:Reference number: S62910
 A:Accession: S63332
 A:Molecule type: DNA
 A:Residues: 1-622 <AER>
 A:Cross-references: EMBL:Z71620; NID:q1302476; PID:e239781; PID:q1302478; MIPS:YNR006w
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: SGD:VPS27; VPT27; VPL23; SSV17
 A:Cross-references: SGD:S0005289; MIPS:YNR006w
 A:Map position: 14R
 C:Function:
 A:Description: controls vacuolar and endocytic traffic through a prevacuolar compartment
 C:Keywords: membrane trafficking

Query Match 2.4%; Score 118.5; DB 2; Length 622;
 Best Local Similarity 20.0%; Pred. No. 2.4;
 Matches 85; Conservative 68; Mismatches 148; Indels 125; Gaps 15;

```

QY 15 GGQTCMLKMSNRIRLTWLSFTLVILVEPPL-----IAHYLTLLDEADGAKR----- 63
Db 196 GGVFCEHSSNSIPPLDLG-----IYEPYVCDSCFEDYDLKRHDSSKSKHRHKK 248
QY 64 ---TFGPRVGNELCEVKNHVDLCRIRESVPELL-----QLEAKROELNSETAKLNK 113
Db 249 KDRDYSTEDEBELIRKAIEISLKESRNSASSEPIVPVESKNEVKROEIEE-EDPDLK 307
QY 114 IEACKSIENAKODLLQLNKVISQTEHSYKEIMQN--OPRLSLPIRLPEKDDAGLPP 170
Db 308 -AAIQESLRBAEAKLR-----SERQKASROMQOPQSPQPIHSDVLSDEKDS----- 357
QY 171 PKATRGCLHNCFDYSCPLTSGEPVYVYDS-----DQFVFGSYLDPLVYKQAFQA 220
Db 358 -----IYMFASLVEKMKSRPLNETLEDKLNLAQRVAFAS 392
QY 221 TARANVYVENADIACLYIIVGEMQEPVYLRPALEKQL-----YSLPHMRTDGHN 272
Db 393 KARLVYALNDRAQKYNLTLEENKISLETIMNTYDRLLQOQISINLSQOYTLPOVPSDPYN 452
QY 273 HVIINLSRKSPTONL--LYNVSTGRAMVAQSTFYTVYRPGFDLVSPLVHAMSEPNFME 330
Db 453 YLTENVQNPAPESYQTPPLQQLSSH-----QYKPOQDVSRQOSVAKANSPT--- 497
QY 331 IPPQVPVKKYLTFFQGEKIESLRS-----SLQEARSFEEEMEGDDPPADYDRIITATLK 384
Db 498 -----TNIDHLKTIDVTPHAQOKPQSHVELAPSDPPYPRKEAEDEGQTQ 540
QY 385 AVQDSK 390
Db 541 AVQDEE 546
  
```

Search completed: September 4, 2002, 14:59:02
 Job time: 121 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 4, 2002, 14:57:00 : Search time 18.3 Seconds
(without alignments)
1226.619 Million cell updates/sec

Title: US-09-809-920-4

Perfect score: 4873

Sequence: 1 MTGYTMLRNGAGNGGQTCM.....DSVLEKTRLPDKTKCFKFI 919

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Archived: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	939	19.3	728	4	US-08-915-337-2
2	115.5	2.4	3075	2	US-08-460-309-5
3	115.5	2.4	3075	2	US-08-125-077-5
4	109.5	2.2	1572	2	US-08-290-731C-5
5	109.5	2.2	1596	3	US-09-356-952-3
6	108	2.2	873	3	US-09-187-331-6
7	108	2.2	873	4	US-09-470-946-6
8	108	2.2	925	2	US-08-392-946-1
9	108	2.2	925	2	US-08-504-169-1
10	108	2.2	925	3	PCT-US94-14893-1
11	103	2.1	2337	6	5455158-1
12	102.5	2.1	829	1	US-07-670-611-2
13	102.5	2.1	829	1	US-08-220-674-2
14	102.5	2.1	829	1	US-08-445-186-2
15	102.5	2.1	829	1	US-08-446-549-2
16	102.5	2.1	829	2	US-08-446-550-2
17	102.5	2.1	1093	4	US-09-315-793-52
18	101.5	2.1	3248	1	US-08-328-254-6
19	101.5	2.1	3248	1	US-08-353-700-1
20	101.5	2.1	3248	5	PCT-US95-16216-1
21	100.5	2.1	800	2	US-08-469-537A-72
22	100.5	2.1	800	2	US-08-469-537A-78
23	100.5	2.1	937	2	US-08-469-537A-105
24	100	2.1	2512	3	US-08-801-263A-9
25	99.5	2.0	428	1	US-09-102-248-9
26	99.5	2.0	428	1	US-08-785-050-2
27	99.5	2.0	428	2	US-08-996-798-2

28	99.5	2.0	655	4	US-09-298-724-2	Sequence 2, Appl1
29	99	2.0	1089	1	US-08-180-195-36	Sequence 36, Appl1
30	99	2.0	1089	1	US-08-477-329-36	Sequence 36, Appl1
31	99	2.0	1089	2	US-08-475-458-36	Sequence 36, Appl1
32	99	2.0	1089	3	US-08-980-400-36	Sequence 36, Appl1
33	99	2.0	1089	4	US-09-583-459A-36	Sequence 36, Appl1
34	99	2.0	1089	4	US-09-583-210-36	Sequence 36, Appl1
35	99	2.0	1089	4	US-09-583-449A-36	Sequence 36, Appl1
36	99	2.0	1579	3	US-08-755-587-184	Sequence 184, App
37	99	2.0	2329	3	US-08-755-587-16	Sequence 16, Appl
38	99	2.0	2386	2	US-09-016-366A-12	Sequence 12, Appl
39	99	2.0	2446	2	US-08-551-356-2	Sequence 2, Appl1
40	99	2.0	2446	5	PCT-US93-12687-2	Sequence 2, Appl1
41	99	2.0	3418	2	US-08-639-501-2	Sequence 2, Appl1
42	99	2.0	3418	3	US-09-044-946-2	Sequence 2, Appl1
43	99	2.0	3418	3	US-09-044-946-2	Sequence 2, Appl1
44	98.5	2.0	1110	1	US-08-118-441-29	Sequence 29, Appl1
45	98.5	2.0	1110	3	US-08-338-579A-29	Sequence 29, Appl1

ALIGNMENTS

RESULT 1
US-08-915-337-2
; Sequence 2, Application US/08915337
; Patent No. 6287802
; GENERAL INFORMATION:
; APPLICANT: Deng, Fan & Xia
; TITLE OR INVENTION: EXT2 Gene
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,337
; FILING DATE: August 21, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: China - 96121928.9
; FILING DATE: October 21, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: ATG50019
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (610) 270-5219
; TELEFAX: (610) 270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 728 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; US-08-915-337-2

Query Match 19.3% Score 939 DB 4 Length 728

Best Local Similarity 31.6%; Pred. No. 1.4e-86;
Matches 257; Conservative 126; Mismatches 256; Indels 174; Gaps 30;

```

QY 140 HSKELMAQONPKSL---PILLEPEKDDAGLP-PPKATRGCRILNCEFDYSCPLTSGP- 194
DB 48 HSISSNDMMNKKRIRDPVVRLP---ADSPIPERGDLSCRMHCCEFYVRC---GFN 99
QY 195 -----PVVYVDSDOFV--FGSYLDPLVKQAFQATARA---NYYVEMNDIACLYILVGE 244
DB 100 PKNKIKVYIYALKKYVDDEGVSVMTISREYIMELMAISDSYTYDDINRACLFPPSDIV 159
QY 245 MOEPVYLPAEIEKOLYSLPHWRIDGHNHVIIN-LSRKSQTONLLYNTSGRAMVAOSTF 303
DB 160 LNQN-TLRIKETQAQMAQLSRW-DGTNHLLENNMLPGGPPDYNTALDVPDRALLAGGEF 217
QY 304 YTVQVRPEFDLVSPVLYAMSEPNMELPPQVPPVKRKLFTFGKIESLNSLOEANSF 363
DB 218 STWYRQGYDVSF-PVYSPLSAE-VDLPEKPGPROYFL-----LSSQVGLHPEY 265
QY 364 EEMEGDPPADYDRRIITLKAVODSKLDQVLYEFTCKNOKPKSLPTMALCGERE--DR 421
DB 266 REDLE-----ALQVKEESVLDKCTNLSEGVLSVR-KRCHKHQPVDY 308
QY 422 LELKLSTFALIITPGDRLYISSGCATRLFEALEVGAVPVLYGEVOLPYQDMLQNMGA 481
DB 309 PVLQEAFTFCVLR--GARL---GQAV-LSDDVLAQAGCPVIADSYILPSEVLDWKRA 361
QY 482 ALVVPKPRVTEVHFLKSLSDSLAMRQ-----GRLEWYFSTADSIFFNTVL--- 531
DB 362 SVVPEEKMSVYSILSIPQRIEEMORQLFMEPVRENN---SAANHQMNSLIWPRE 417
QY 532 ----AMITRIQIIPAAPFIREAAAEIPHRSGKAAGTDPNMADNDLDGPVETEPYPSP 587
DB 418 QMDQIINDRIYPPAALSYEE-----WNPDPAY-- 445
QY 588 RYLNFTLVTDYFSMNACAPRHPHPDPVLPSEAKFLSGTGFRPIGGAGSG 647
DB 446 -----KMGVSNPFLF-----PLIPPOS----- 463
QY 648 KEFOALGQNVPRROFTVMTYFREEVYAMSLERLNGLPKLVVVVVVWNSP-KLPSEDL 706
DB 464 -----OGFPAIVLTVDYVESLFEVITEVSKVPSLSKLIVVWNNQNNPDEDS 510
QY 707 LMPDIGVPIWVTEKSNLNRFLPMNEIETALISIDD-AHLRHEIMGEFVWREAR 765
DB 511 LMPRIAPLKVYRAEKLSRFFPYDEIEEAVLAIDDDITMLTSELQGYVWREPP 570
QY 766 DRIYGFGRYHAMDIPHQSMLYSNSYCELSMWLTGAFFHKYVYLYSVYVMPQAIRDMY 825
DB 571 DRLVGYDRLHLMDEHKNKKWKESEMTEWNSMVLGAFFHKYFNLYLYKKMPGDIDKNW 630
QY 826 DEVINCDDIMNPLVSHITKRPKYTSKMTFRCPCGP--QALSHDSSHFERHKCINFP 883
DB 631 DAHNNCDIAMNPLVANVTGAKAVIKYPRKKFKPECTAIDGLSDQTHWERSECKINF 690
QY 884 VKYVGYMPLLYTOFRVDSVLFKTRLPDHDKCF 916
DB 691 ASVFGTMPLKLVHRADPVLYIKDDFP-EKLKSF 722

```

RESULT 2
US-08-460-309-5
Sequence 5, Application US/08460309
Patent No. 5837496
GENERAL INFORMATION:
APPLICANT: Engvall, Eva
APPLICANT: Leivo, Ilmo
TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
TITLE OF INVENTION: Fragments and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700

```

CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC Compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460.309
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125.077
FILING DATE: 22-SEP-1993
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472.319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919.951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3075 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-460-309-5

```

Query Match 2.4%; Score 115.5; DB 2; Length 3075;
Best Local Similarity 20.1%; Pred. No. 0.27;
Matches 120; Conservative 102; Mismatches 237; Indels 139; Gaps 27;

```

QY 71 NELCEVHAYLD-LCRIRSEVSEBLLQLEPAKQELINSEIAKINKLECKSSTENAKODLL 129
DB 1717 NATLELKAEDLLSQIOENYQKPLEELEVLKEASHYLSKHNHLEKAALVREAAKMQ 1776
QY 130 QLRKVI-----SQTHSHYKELMAQONPKSLPILLEPEKDDAGLP-PPKATRGCRILNCEFD 184
DB 1777 ESNHLLMVANNLREFSDFDKLHVQEQNLTSSELIV---OGRGILDAAAAOQDAVODALE 1832
QY 185 YSRCLPISGFPVYVYDSDOFVFGS-----YLDPLVKQAFQATARANYVTEN--ADIACL 237
DB 1833 HLE-----DHDDKLLMSAKIRHNIHIDLVMHMSORNAVLDYRAEDHATEFORL 1881
QY 238 YVILVGEKQVYVLPRAELKOLYSLPHWRTDG---HNHVIINLSRKSQTONLLYN--V 291
DB 1882 ADVLYSGLEN---IRNSLNTASNAAYVHYNQSLIESESEELADAHRTVETSLSSSLV 1938
QY 292 STGRAMVAOSTFTYVQVR-----PGFDLYVSPVLYAHMS--EPNEMETIPQ----- 334
DB 1939 SNCKAAVQRRSSRLKESNNLSRLPGIALSELRLNTRNFQSNVAEITRQTHESLILIR 1998
QY 335 -VP-----VKRKYLTFFQGEKIESLSLSLOEARSPFEEMEGODPPADYDDRIIATLK-- 384
DB 1999 AIPBGIRDKAKTKRELAT---SASQSAVSTLRVYAGISQDLMTWSAS--LSRVNTTLRET 2053
QY 385 -AVQDSKLDQVLYEFTCKN--QPKRPSLPTMALCGEREDRLKLSTFALIITPGDP 439
DB 2054 HOLQDSYMATILLAGRKVKDVEIQAKVLF-----DRKPLKMLEENSRNLSL 2102
QY 440 RLVISSGCATRLFEALEVGAVPVLYGEVOLPYQDML---QNMNLAALVVPKPRVTEVHFL 496

```



```

Db      2103 KLLISQ--ARKQAASIKVA---VSADROCIIRAYQPOISSTNYNTLTLLNKTKTQEPDNLFFY 2157
Oy      497 LRSLSDDLLAMR-RQGR--FLWEYTFSTADISFTNTVLAMITRQIQIAPAPRREAAFI 553
Db      2158 LGSSTASDFLAEMRRGRAFLMD-----LGSSTRLFEPDPPIIDNRWMSI 2204
Oy      554 -----PHRSGKAGTDPNMADNGDDL-----GVETEP 582
Db      2205 HVARGNIGSLSVKEMSSNQKSPYTKSRKSPGT-ANVLVDNNSTLMPVGLGGQIKKSP 2261

RESULT      3
US-08-125-077-5
; Sequence 5, Application US/08125077
; Patent No. 5872231
; Patent No. 5872231 5840863
; GENERAL INFORMATION:
; APPLICANT: Engvall, Eva
; APPLICANT: Leivo, Ilmo
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
; TITLE OF INVENTION: Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/125,077
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US 94/10730
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/472,319
; FILING DATE: 30-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/919,951
; FILING DATE: 27-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9721
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3075 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-125-077-5

Query Match      2.4%; Score 115.5; DB 2; Length 3075;
Best Local Similarity 20.1%; Pred. No. 0.27;
Matches 120; Conservative 102; Mismatches 237; Indels 139; Gaps 27.

Oy      71 NELCEVKAHLD-LCRIRSEVSEELLQLEAKROELNSEIAKLNLKTEACKKSTENAKODLL 129
Db      1717 NATLELKAEDLDSIOEYOKRPLELEVLKAEASVLSKHNNELKAABALVREAAKMQ 1776
Oy      130 QAKNI-----SQTEHSYELMAQNPKSLPIRLLPEKDDAGLPPPKATYTRCRLHNCED 184
Db      1777 ESNHLLWNAVLREPSDKKLHVQEOUNTTSELV---QGRGLIDAAQAQDAVDAVDAL 1832

```

```

01 185 YSRPRLTSGFVVYVYDDQCVFSS-----YLDYLVQCAQCAAFARAVYTEN--ADICL 237
02 1833 HLE-----DHQDKLLMSAKIRHHIDDLVNMHMQRRNAVLDLYAEDHAFEPRL 1881
03 238 YVILVGEQEVVLPAELEKQLYSLPHWTFD--HNHYIINLSKSDPTQMLYNN--V 291
04 1882 ADVYISGLE---IRNVSLNATSAAYVHHYNIQSLIEESEELARDHARVETSLSESLV 1938
05 292 STGAMAAQSTFYVQVR-----PGFDLYVSPVHAMS--EPNFEIPEQ----- 334
06 1939 SNGRAAVALORSRFLKEBNLSRLPLGIALSELRLKRTKRFQENAEVITRQNESLILR 1998
07 335 -VP-----VKRKYETFGQEKIESLRSLQARSFEEEMEGDPPADDRIITLK-- 384
08 1999 AIPGIDKDKAKTTELAT---SASQASVSTLRVAYALSOELINTAS--LSRVNTTLKET 2055
09 385 --AVQDSKLDQVLVEFTCKN--QPKPSLPTMALCGEREDRLLELIKSTFALLITPPDP 439
10 2054 HOLLDSTMTATLLLAGRKVKVDVEIOAKVLF-----DRLKPLKMEMLSHNLSEI 2102
11 440 RLVISSGATRLFEALBVGAVPVVLGQVQDLPYQDML---QNNKALVYPRKRVYEVFL 496
12 2103 KLLISQ--ARKQASIKVA---VSADBCIRAYQPOISSTNTVTLTLNKTQEPDNLIFY 2157
13 497 LRSLSDSLAMR--RGR--FLMETVFSTADSIFFNTVLAMIRTRIQIPAPFREEAAEI 553
14 2158 LGSSTADGFLAVEKRRRVAFLWD-----LQSGSRLEFPFPITDDNRHMSI 2204
15 554 -----PHRSKAGATDPNNMADGDDL-----GVPETEP 582
16 2205 HVARFGNISLVSKESSNOKSPFKTSKSPGT-ANVLVNNSTLMEVGLGQIKKSP 2261

RESULT 4
US-08-290-731C-5
Sequence 5, Application US/08290731C
Patent No. 5843646
GENERAL INFORMATION:
APPLICANT: BOWTELL, David Douglas Lawrence
TITLE OF INVENTION: DNA MOLECULES ENCODING MURINE
TITLE OF INVENTION: SON OF SEVENLESS (MSOS) GENE,
TITLE OF INVENTION: AND MSOS POLYPEPTIDES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: SUGHRUE, MTON, ZINN, MACPEAK & SEAS
STREET: 2100 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/290,731C
FILING DATE: 17-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU93/00068
FILING DATE: 17-FEB-1993
PRIOR APPLICATION DATA: PLO921/92
APPLICATION NUMBER: PLO921/92
FILING DATE: 17-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: KIT, Gordon
REGISTRATION NUMBER: 30,764
REFERENCE/DOCKET NUMBER: 0-36066
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860

```

TELEX: 6491103
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1572 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-290-731C-5

Query Match 2.2%; Score 109.5; DB 2; Length 1572;
 Best Local Similarity 19.6%; Pred. No. 0.33; Mismatches 359; Indels 313; Gaps 54;
 Matches 198; Conservative 142; Indels 359; Gaps 54;

QY 76 VKHVLDCRIRSEVSEELLQLEAKROELNSELAKNLKIEA--CKSIENAKQDLDLQKN 133
 DB LKKVLEQVHPRYTAKEDALLY-----VEKLCRLMLAKKPLPHSVQDVEEKNV 126
 QY 134 -----VISQTEHSYKELMAQONQPKLSLP--IRLLPEKDDAGLPPEKATRGCRRLHNCFD 184
 DB 127 KSPFAPIDQMALNKAKEVINSKKRSVLPTEKVHTLLQKD-----VLQ 169
 QY 185 YSRCPPLSGFPVYVYSDQFVFGSYL---DPLVKAQAFATARANYVTENADIACLIV- 239
 DB 170 YKIDSSVSAGFLVAVL---EYISADILKMAQGVYIKIAHCETIKEDIEVMMADRVLMQDL 226
 QY 240 -----ILVGMQEPVVLPAELEKOLYSLPH---WRTGHNHVIINLSRK-----SD 283
 DB 227 NQSEAHILPSPISPAPORASATYEETVELIHDEKQYRD--LHMIVFREELVKIYSD 284
 QY 284 TQNL-----LYNVS-----TGRAM--VAQSTFYVQYRPG 311
 DB 285 PRELEPFSNIMDIYEVTYVLLGLSLIEDVIMSQEOSAPCVSGFEELAEAEFDYKKYA 344
 QY 312 FDLYV---SPLVHAMESEPNMEIP-----PQ---VPVKRYLFTFGGEKI 350
 DB 345 YDVTQSQRDALNNLSKPGASSLTAGHGRDAVKYLLPRLLVPLICHAFFYDYIKHL 404
 QY 351 ESLSLSQLEARSFEEMEGDPPADYD--DRIATLKAQVDSKLD-----QVLVEFTCKNQP 404
 DB 405 KDLSSSDDIIESFOVGLLPHLCHDLEKVMASLSKERQVPSGRVRQOLAIERTRELQ 464
 QY 405 KPSLPTEM-----ALCGE--REDRLLEL-----KLSFF--ALITPGDPLRIVS 444
 DB 465 KVE---HMEKDVQGNQNEFIREDSLSKSGKRIMSEKRVFLPDGLMVLCAKNTKQTP 521
 QY 445 SGCAT-----RLFEALEVGAAPV-----VLGEQVQLPY-----QDMLQWNE 480
 DB 522 SAGATADYRLKEKYMRRYVDINDRPSDDLKNSFELAPRMQPTIVLAKNAQKHQWMA 581
 QY 481 AALVYPRPRVTEVHF--LMSLSDSDLLAMRQGRFLMEYI--ESTADISINFTVLAITRTR 537
 DB 582 DLMVITKSMIDRLDSIIDLIERKHLRMPSP-----EYKFAVPVSGQNIIVL----- 630
 QY 538 IOIPADIREEAAEIPHRSGKAAGTDPNMAANDGDLDPVETEPAPASPYRLNFTLV 597
 DB 631 -----EBRESAGVIMIGATL-----CKLIERLYHIYADPFEVVFTEFLT- 670
 QY 598 TDYRSMNCARPGPHLEPHTPDEVLPSSEAKFLGSGTGFPRIIGGAGSGSKEFOALG- 656
 DB 671 ---YR-YFCSPQDLDLIVFRFNIPDS-LVYQDTGT-----AGAGGMC-----GVGCD 714
 QY 657 ---NPRE-----QFTYV-----MLTYERE-----EVLMSLEKLNG 685
 DB 715 KEKHNHREDMKRYRREYQVQFRVNLVLRHVVHHPYFEKDPMLLEKLLNFLEHVG 774
 QY 686 LPLINKVVVVVNSPKLPSSEDLMPDGVPLMVYVTEKNSLNRF-----PNNLEIETEILS 742
 DB 775 -KSMRK-----WDSYVLKIVQKKNDEKSNKKTIVAYVGHDPPTIEHLS 817
 QY 743 IDDD-----AHLRHDEIMFGFRVWRAR--DRIYGFGRYHANDIPQSWL-----YN 788
 DB 818 VPNDEITLLTLHLPLLELAQUTLLLEFEMKYNKIVSELVGSWTKKDKVEKSPNLKIKMKT 877

QY 789 SNVSCELSMVLGAFFHKKYAYLYSYVPOAIRDMVDEYINCEDIAMNFLTSHTRKP 848
 DB 878 TWTWMIKESITAEADYERELA-----IMQALIEVMM---VLEELNNGILSIYAANGT 929
 QY 849 IKV-TSRWTFRCGCPQ-----ALSHDD-SHFERRK-----CINFP 883
 DB 930 ASVYRLRWTFQ--GLPERYRKFLRECRLELSDHLLKKYQERLRSINPCVPFF 979

RESULT 5
 US-09-356-952-3
 ; Sequence 3, Application US/09356952
 ; Patent No. 6117663
 ; GENERAL INFORMATION:
 ; APPLICANT: Boriack-Stodin, Ann.
 ; APPLICANT: Margalit, S. M.
 ; APPLICANT: Bor-Sogli, Dafna
 ; APPLICANT: Cole, Philip
 ; APPLICANT: Kuriyan, John
 ; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
 ; TITLE OF INVENTION: THEROOF
 ; FTE REFERENCE: 600-1-228N
 ; CURRENT APPLICATION NUMBER: US/09/356,952
 ; CURRENT FILING DATE: 1999-07-19
 ; EARLIER APPLICATION NUMBER: 60/093,631
 ; EARLIER FILING DATE: 1998-07-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: patentin Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 1596
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 ; US-09-356-952-3

Query Match 2.2%; Score 109.5; DB 3; Length 1596;
 Best Local Similarity 19.6%; Pred. No. 0.34;
 Matches 198; Conservative 142; Mismatches 359; Indels 313; Gaps 54;

QY 76 VKHVLDCRIRSEVSEELLQLEAKROELNSELAKNLKIEA--CKSIENAKQDLDLQKN 133
 DB 77 LKKVLEQVHPRYTAKEDALLY-----VEKLCRLMLAKKPLPHSVQDVEEKNV 126
 QY 134 -----VISQTEHSYKELMAQONQPKLSLP--IRLLPEKDDAGLPPEKATRGCRRLHNCFD 184
 DB 127 KSPFAPIDQMALNKAKEVINSKKRSVLPTEKVHTLLQKD-----VLQ 169
 QY 185 YSRCPPLSGFPVYVYSDQFVFGSYL---DPLVKAQAFATARANYVTENADIACLIV- 239
 DB 170 YKIDSSVSAGFLVAVL---EYISADILKMAQGVYIKIAHCETIKEDIEVMMADRVLMQDL 226
 QY 240 -----ILVGMQEPVVLPAELEKOLYSLPH---WRTGHNHVIINLSRK-----SD 283
 DB 227 NQSEAHILPSPISPAPORASATYEETVELIHDEKQYRD--LHMIVFREELVKIYSD 284
 QY 284 TQNL-----LYNVS-----TGRAM--VAQSTFYVQYRPG 311
 DB 285 PRELEPFSNIMDIYEVTYVLLGLSLIEDVIMSQEOSAPCVSGFEELAEAEFDYKKYA 344
 QY 312 FDLYV---SPLVHAMESEPNMEIP-----PQ---VPVKRYLFTFGGEKI 350
 DB 345 YDVTQSQRDALNNLSKPGASSLTAGHGRDAVKYLLPRLLVPLICHAFFYDYIKHL 404
 QY 351 ESLSLSQLEARSFEEMEGDPPADYD--DRIATLKAQVDSKLD-----QVLVEFTCKNQP 404
 DB 405 KDLSSSDDIIESFOVGLLPHLCHDLEKVMASLSKERQVPSGRVRQOLAIERTRELQ 464
 QY 405 KPSLPTEM-----ALCGE--REDRLLEL-----KLSFF--ALITPGDPLRIVS 444
 DB 465 KVE---HMEKDVQGNQNEFIREDSLSKSGKRIMSEKRVFLPDGLMVLCAKNTKQTP 521
 QY 445 SGCAT-----RLFEALEVGAAPV-----VLGEQVQLPY-----QDMLQWNE 480

```

Db 522 SAGATADYDILKEKRYRRARDINDRPSDDLKNSFELAPQRPYVLTAKNAQKHMMMA 581
OY 481 AALVPRKRVTEVHF--LRSLSOSDILAMRQGRFLMETY--ESTADSIENVYLAIRTR 537
Db 582 DLLMVTIKSMILDRILDSILQDIERKHPLRMSP-----EIKFAVPDSGNIVL----- 630
OY 538 IQIPAIRREMAAEIHRSGKAAGTDPNMADNDLDLGVETEPYPASRYLRNFTLV 597
Db 631 -----EERESAGVMIKATL-----CKLIRLYHIADPTFVTFLLT- 670
OY 598 TDFRSWNCAPRPHLPPTDPVLPSEAKFLGSGTFRPIGGAGSGKEFOALG- 656
Db 671 ---YR-YEGSPQDLQLLVERFNTDPS-LYQDTGT-----AGACGCG-----GYGCD 714
OY 657 ---NPRE-----OFTVV-----MLTYERE-----EYLMNSLERING 685
Db 715 KEHNSHREDMKRYRKYEQVQVQFRLVNLRYHWYDHFYDEKDPMLLEKLNLEHNG 774
686 LPYLKRVVWVNSPKLPSEDLMPDVGPIMVVTEKNSLNRLP---PNNETETALIS 742
Db 775 -KSMRK-----WDSVLKTYORKNEQKSNKIYVAYGHDPPRIEHLIS 817
OY 743 IDDD-----AHLRHEIMGFRVYRREAR--DRIVGFRGYHAMDIPQOSML-----YN 788
Db 818 VPNDITLLTLHLPLRLAQLTLLEFEMTKVVKPSSELVGSFWTKKDKVEKSPMLIKTKHT 877
OY 789 SNVSCELSNVLTGAAPFKYAYLYSVMPQAIRDNDVDEXINCEDIAMNPLVSHITRPP 848
Db 878 TWTWRIEKSITELEANEERLA-----IMQRALEVM---VMLENNFNGILSIYAAMGT 929
OY 849 IKV-TSMWTRCQCPQ-----ALSHDD-SHFHEHKK-----CINEF 883
Db 930 ASVYRLMWTQ--GLPERYRKFLECRELSDDLHKYOEKRLRSINPCVPFF 979

RESULT 6
US-09-187-331-6
; Sequence 6, Application US/09187331
; Patent No. 6043056
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Gorgone, Gina A.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS
; FILE REFERENCE: PF-0631 US
; CURRENT APPLICATION NUMBER: US/09/187,331
; CURRENT FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 873
; TYPE: PRP
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: g189650
US-09-187-331-6

```

Query Match 2.28; Score 108; DB 3; Length 873;
 Best Local Similarity 23.8%; Pred. No. 0.17;
 Matches 89; Conservative 48; Mismatches 127; Indels 110; Gaps 23;

```

OY 541 PAAPIREMAAEIPHR--SGKAAGTDPNMADNDGLDGPVETEPYPASRYLRNFTLV 598
Db 401 PAARLR--PSDVPDKYSTRYEGSIARNLSR-----EPNQHFKPYLKHFLPKRL 447
OY 599 DFYRSWNCAPRPHLPPTDPVLPSEAKFLGSGTFRPIGGAGSGKEFOALGNV 658
Db 448 HFAKSDRIEPLTFYLDQWQL-ALNPSEKRYCGSGF-----HGSNDVFSMQALFVGYG 500

```

```

OY 659 PROFTVVMLTVEERY--LMNSLERL-----NG-----LPYLKRVVWVNSPKLPSEDL 707
Db 501 PGRHGIADTFENIEYVNLMLCDLNLTPAPNNGTSGLSLHNLKNPY--TPKHPE--V 556
OY 708 WPDIGVIMVVRTEKNSL-----NNRFLPMNEIETETALISIDDAHLRHEIMGFRVWRE 763
Db 557 HPLVQCFP--TRNPRODLGSCNPSILPIEDPQTQFNLVAAEKIINHETLPYG----- 608
OY 764 ARDRIVGFRGYHAMDIPHOSWLYNSYSCELSNVLTGAAPFKKYA--YLYSVMPQAIR 822
Db 609 -RRRV-----LQKENTICLLSQ-----HQFMSGYSQDILMLPMTS 642
OY 823 DMYD--EYINCEDIAMNPLVSHITRKPPIKVTSMWTRCQCPQALSHDSSHFERHKCI 880
Db 643 YTVDRNDSFSTEDFS-NCLYQ-----DRIPLSPV-----HKC- 674
OY 881 NFF---YKV-YGYM 890
Db 675 SFYKNNTKVSYGFL 688

```

```

RESULT 7
US-09-470-946-6
; Sequence 6, Application US/09470946
; Patent No. 6358923
; GENERAL INFORMATION:
; APPLICANT: Yue, Henry
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Gorgone, Gina A.
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS
; FILE REFERENCE: PF-0631 US
; CURRENT APPLICATION NUMBER: US/09/470,946
; CURRENT FILING DATE: 1999-12-22
; EARLIER APPLICATION NUMBER: US 09/187,331
; EARLIER FILING DATE: 1998-11-06
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 873
; TYPE: PRP
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: g189650
US-09-470-946-6

```

Query Match 2.28; Score 108; DB 4; Length 873;
 Best Local Similarity 23.8%; Pred. No. 0.17;
 Matches 89; Conservative 48; Mismatches 127; Indels 110; Gaps 23;

```

OY 541 PAAPIREMAAEIPHR--SGKAAGTDPNMADNDGLDGPVETEPYPASRYLRNFTLV 598
Db 401 PAARLR--PSDVPDKYSTRYEGSIARNLSR-----EPNQHFKPYLKHFLPKRL 447
OY 599 DFYRSWNCAPRPHLPPTDPVLPSEAKFLGSGTFRPIGGAGSGKEFOALGNV 658
Db 448 HFAKSDRIEPLTFYLDQWQL-ALNPSEKRYCGSGF-----HGSNDVFSMQALFVGYG 500
OY 659 PROFTVVMLTVEERY--LMNSLERL-----NG-----LPYLKRVVWVNSPKLPSEDL 707
Db 501 PGRHGIADTFENIEYVNLMLCDLNLTPAPNNGTSGLSLHNLKNPY--TPKHPE--V 556
OY 708 WPDIGVIMVVRTEKNSL-----NNRFLPMNEIETETALISIDDAHLRHEIMGFRVWRE 763
Db 557 HPLVQCFP--TRNPRODLGSCNPSILPIEDPQTQFNLVAAEKIINHETLPYG----- 608
OY 764 ARDRIVGFRGYHAMDIPHOSWLYNSYSCELSNVLTGAAPFKKYA--YLYSVMPQAIR 822
Db 609 -RRRV-----LQKENTICLLSQ-----HQFMSGYSQDILMLPMTS 642
OY 823 DMYD--EYINCEDIAMNPLVSHITRKPPIKVTSMWTRCQCPQALSHDSSHFERHKCI 880

```

Db 643 YVVDNDSTEDFS-NCILYQ-----DRIPLSPV-----HRC- 674
QY 881 NFE---VKV-YGYM 890
Db 675 SFYKNNTKVSYGFL 688

RESULT 8

US-08-392-946-1
; Sequence 1, Application US/08392946
; Patent No. 5939269
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: The Regents of the University of California
; APPLICANT: Goldfine, Ira D.
; APPLICANT: Grupe, Andrew
; APPLICANT: Maddux, Betty A.
; APPLICANT: Spencer, Steven
; APPLICANT: Stewart, Timothy A.
; TITLE OF INVENTION: Antagonists to Insulin Receptor Tyrosine
; TITLE OF INVENTION: Kinase Inhibitor
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/392,946
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/182241
; FILING DATE: 14-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Kudinec, Jeffrey S.
; REGISTRATION NUMBER: 36,575
; REFERENCE/DOCKET NUMBER: P0875P1PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-8228
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 925 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-392-946-1

Query Match 2.2%; Score 108; DB 2; Length 925;
Best Local Similarity 23.8%; Pred. No. 0.18;
Matches 89; Conservative 48; Mismatches 127; Indels 110; Gaps 23;

QY 541 PAAPIREAAAEIPIHR--SGKAAGTDPNADNGLDLAGVETPEPIYASPRYLNRFTLTVT 598
Db 453 PAALRL--PSDVPDKYYSFNYEGIAIRNLSCR-----EPNQHFKPYLKHFLPKRL 499
QY 599 DFRSNNACGAPFLHFTEDPVLPEAKFLGSGTGFRPIGGAGSGSGKEFOALGCV 658
Db 500 HFAKSDIEPLTLTDQOML-ALNPSEKRYCGSGF-----HESDVFSSMQALEVGYG 552
QY 659 PDEQTVMLTYEREV--LANSLERL-----NG---LPYLNKVVVWMSPKLPSEDL 707
Db 553 PGRKHGIEADTFENIEVYNNLMCDLNLTPARNNGTHOSLNLHLLKNPY--TPKHPRK--V 608

QY 708 WPDIGVPIVNVTEKNSL-----NNRFLPWNIEIEAIIISIDDAHLRHDEIMFGRWRE 763
Db 609 HPLVOCPE--TRNPRDNLGSCNPSILPIEDFOFNLIVAEKIKHETLPYG----- 660
QY 764 ARDRIVGFGRRYHANDIPHQSLYNSNYCELSMWLTGAFFHKYYA-YLYGVVPOAIR 822
Db 661 -RPRV-----LOKENTICLSQ-----HOFMSGVSODILMPLWTS 694
QY 823 DMYD--EYINCEDIAMNPLVSHITRKPPIKYSRWTFRCPGCPQALSHDHSFHERKCI 880
Db 695 YVVDNDSTEDFS-NCILYQ-----DRIPLSPV-----HRC- 726
QY 881 NFE---VKV-YGYM 890
Db 727 SFYKNNTKVSYGFL 740

RESULT 9

US-08-504-169-1
; Sequence 1, Application US/08504169
; Patent No. 5968508
; GENERAL INFORMATION:
; APPLICANT: Goldfine, Ira
; APPLICANT: Grupe, Andrew
; APPLICANT: Henzel, William
; APPLICANT: Maddox, Betty
; APPLICANT: Spencer, Steven
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Harrison, Denise M.
; TITLE OF INVENTION: Antibodies to Insulin Receptor Tyrosine Kinase Activation I
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/504,169
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/14893
; FILING DATE: 28-Dec-1994
; PRIOR APPLICATION NUMBER: 08/182241
; APPLICATION NUMBER: 08-Jan-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kudinec, Jeffrey S.
; REGISTRATION NUMBER: 36,575
; REFERENCE/DOCKET NUMBER: P0875P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-8228
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 925 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-504-169-1

Query Match 2.2%; Score 108; DB 2; Length 925;
Best Local Similarity 23.8%; Pred. No. 0.18;
Matches 89; Conservative 48; Mismatches 127; Indels 110; Gaps 23;

[illegible]

```

10 RESULT
11 PCT-US94-14893-1
12 ; Sequence 1, Application PC/TUS9414893
13 ; GENERAL INFORMATION:
14 ; APPLICANT: Genentech, Inc.
15 ; APPLICANT: The Regents of the University of California
16 ; APPLICANT: Goldfine, Ira D.
17 ; APPLICANT: Grube, Andrew
18 ; APPLICANT: Maddux, Betty A.
19 ; APPLICANT: Spencer, Steven
20 ; APPLICANT: Stewart, Timothy A.
21 ; TITLE OF INVENTION: Antagonists to Insulin Receptor Tyrosine Kinase Inhibitor
22 ; NUMBER OF SEQUENCES: 1
23 ; CORRESPONDENCE ADDRESS:
24 ; ADDRESSEE: Genentech, Inc.
25 ; STREET: 460 Point San Bruno Blvd
26 ; CITY: South San Francisco
27 ; STATE: California
28 ; COUNTRY: USA
29 ; ZIP: 94080
30 ; COMPUTER READABLE FORM:
31 ; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
32 ; COMPUTER: IBM PC compatible
33 ; OPERATING SYSTEM: PC-DOS/MS-DOS
34 ; SOFTWARE: palin (Genentech)
35 ; CURRENT APPLICATION DATA:
36 ; APPLICATION NUMBER: PCT/US94/14893
37 ; FILING DATE:
38 ; CLASSIFICATION:
39 ; PRIOR APPLICATION DATA:
40 ; APPLICATION NUMBER: 08/182241
41 ; FILING DATE: 14-JAN-1994
42 ; ATTORNEY/AGENT INFORMATION:
43 ; NAME: Kubinec, Jeffrey S.
44 ; REGISTRATION NUMBER: 36,575
45 ; REFERENCE/DOCKET NUMBER: 875P1PCT
46 ; TELECOMMUNICATION INFORMATION:
47 ; TELEPHONE:
48 ; TELEFAX: 415/952-9881
49 ; TELEX: 910/371-7168
50 ; INFORMATION FOR SEQ ID NO: 1:
51 ; SEQUENCE CHARACTERISTICS:
52 ; LENGTH: 925 amino acids
53 ; TYPE: amino acid
54 ;

```

```

;          TOPOLOGY: linear
PCT-US94-14893-1

```

Query Match	2.2%	Score 108;	DB 5;	Length 925;
Best Local Similarity	23.8%	Pred. No. 0.18;		
Matches	89;	Conservative	48;	Mismatches 127;
			Indels	110;
			Gaps	23;

[illegible]

```

RESULT 11
5455158-1
Patent No. 5455158
APPLICANT: VOGEL, TIKVA; LEVANON, AVIGDOR; WERNER, MOSHE M.;
GUY, RACHEL; PANEI, AMOS
TITLE OF INVENTION: FIBRIN BINDING DOMAIN POLYPEPTIDES AND
USES AND METHODS OF PRODUCING SAME
NUMBER OF SEQUENCES: 20
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/58,241
FILING DATE: 04-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 526,397
FILING DATE: 21-MAY-1990
APPLICATION NUMBER: 345,952
FILING DATE: 28-APR-1989
APPLICATION NUMBER: 291,951
FILING DATE: 29-DEC-1986
SEQ ID NO:1:
LENGTH: 2327
5455158-1

```

Query Match 2.1%; Score 103; DB 6; Length 2327;
Best Local Similarity 22.2%; Pred. No. 3.1;
Matches 119; Conservative 67; Mismatches 184; Indels 166; Gaps 32;

Q7	25	ENQYDLESTATSVNI	PDLPGRKY	-LVNYQJISEDBQSJLST	SYQTAPADAPDPYD	7922
Db	734	EPQYDLESTATSVNI	PDLPGRKY	-LVNYQJISEDBQSJLST	SYQTAPADAPDPYD	7922
QY	300	--QSFYVVOX-RP----	GFDLVSPVLVHMS-	EPNMEIPQVPAKR-----	KYLETE	3455
Db	793	QYDQJSTVYKRSRQAP	PTTGRTYYSB	VESSSTELNLPETANSVTL	SDQPGQVNIIT	8522
QY	346	QG--EKIESLSSLOEANS	FEENEGOPADYDR	IIATLKAODSKLDQVL-	VEFTCKN	4020

Db 853 YAVEENQESTPVVIO-----OETGTGPRSD-----TVPSPRDLQFVEVTVKTYIMW 899
QY 403 QPKPSLPTBMA-----LOGEREDLELLKSTFALI--ITPGDRLVISSCATRLPE 453
Db 900 TPPEASVITGRVDVIVNLPGEHGRLPISR-NTFAEVTGLSPG-----VYYFRK 948
QY 454 ALEV-----AVPVLGEQVQLPYQDMLQW-NEAALVY-----PKPRVTEVHFLRLSLD 502
Db 949 VFAVSHGRSKPLTQOTTKLAPLTLQFVNEIDSTVLYVRTPPRAQITGYL----- 1001
QY 503 SOLLMARQGRFLWETFEYSTADSIENVTIAMIRTRIQIPAPRIEBAAEIPHRSKAAG 562
Db 1002 --TVGLTRRGQ--PROYVWGSPVVKYPLRLNQPASEYTVSLVIAKNGQESP----KATG 1052
QY 563 TDPNNADNDLDLGPVEPEPPASPRYLNFLLTYTDFRSNMCACGPHLPHPPDPV 622
Db 1053 VFTT-----LQPGSSILPPY-----NTEVTESTIVITWTPAPR----- 1084
QY 623 LPSEAKFLGSGTGFRPIGGAGSGKEFOALGNVPRE-----QFTVVMLTRE--- 673
Db 1085 -----IGFKLGVPR-----SOGGAPREVTSDSGSIYVSGILTPGVEYVY 1123
QY 674 --EVLMSLERLNGLPYLNKVVVYVNSKLPSEDL--LMPDIGVPINVTREKNS 724
Db 1124 TIQVLRDQER--DAPIVNVKVV--TLPSPPTNHLBANPDTGY--LTVSWERST 1171

RESULT 12

US-07-670-611-2
; Sequence 2, Application US/07670611
; Patent No. 5330892
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: White, Raymond
; APPLICANT: Nakamura, Yusuke
; TITLE OF INVENTION: Gene Mutated in Colorectal Cancer of
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch et al.
; STREET: 1001 G Street
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20001-4597
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/670,611
; FILING DATE: 19910313
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.33981
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 829 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens

US-07-670-611-2

Query Match 2.1%; Score 102.5; DB 1; Length 829;
Best Local Similarity 29.4%; Pred. No. 0.55;
Matches 30; Conservative 22; Mismatches 43; Indels 7; Gaps 3;

QY 59 EAGKRIFGPRVGNELCEVKNHVDLCRIRSVSEE-LLOLEARKROELNSEIARKLNKIEAC 117
Db 43 EREDRLLEKRLAKAOEQSH--LMREHEDVOERTTLREERITELSHVIAELNKKIDRL 99
QY 118 KKSIEAKODLLQKNVISTQTESYKE--LQAQPKISLP 156
Db 100 QGTTREDEYSSELNSELISOQHEVNEDSRSMDOOTSISP 141

RESULT 13

US-08-220-674-2
; Sequence 2, Application US/08220674
; Patent No. 5571905
; GENERAL INFORMATION:
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: White, Raymond
; APPLICANT: Nakamura, Yusuke
; TITLE OF INVENTION: Gene Mutated in Colorectal Cancer of
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch et al.
; STREET: 1001 G Street
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20001-4597
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/220,674
; FILING DATE: 31-MAR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/670,611
; FILING DATE: 13-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.33981
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 829 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-08-220-674-2

Query Match 2.1%; Score 102.5; DB 1; Length 829;
Best Local Similarity 29.4%; Pred. No. 0.55;
Matches 30; Conservative 22; Mismatches 43; Indels 7; Gaps 3;
QY 59 EAGKRIFGPRVGNELCEVKNHVDLCRIRSVSEE-LLOLEARKROELNSEIARKLNKIEAC 117

Db 43 EREDDLEKLAQAQCSH---LMREHEDVQERTTLRYEBRITELSHVIAELNKKIDRL 99
 QY 118 KKSIEAKODLQKLVISQTEHSYKE---LMAONQPKLSLP 156
 Db 100 QGTTIREDEYSELRSLSOSQHEVNEDESRMDDQTSVSLP 141

RESULT 14

US-08-445-186-2
 ; Sequence 2, Application US/08445186
 ; Patent No. 5576422

GENERAL INFORMATION:

APPLICANT: Vogelstein, Bert
 APPLICANT: Kinzler, Kenneth W.
 APPLICANT: White, Raymond
 APPLICANT: Nakamura, Yusuke
 TITLE OF INVENTION: Gene Mutated in Colorectal Cancer of
 TITLE OF INVENTION: Humans
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner, Birch et al.
 STREET: 1001 G Street
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20001-4597

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/445,186
 FILING DATE: 19-MAY-1995
 CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/220,674
 FILING DATE: 31-MAR-1994
 APPLICATION NUMBER: US 07/670,611
 FILING DATE: 13-MAR-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.33981

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 829 amino acids
 TYPE: amino acid
 STRANDEDNESS: single

TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ANTI-SENSE: NO

ORIGINAL SOURCE:
 ORGANISM: Homo sapiens

US-08-445-186-2

Query Match 2.1%; Score 102.5; DB 1; Length 829;
 Best Local Similarity 29.4%; Pred. No. 0.55;
 Matches 30; Conservative 22; Mismatches 43; Indels 7; Gaps 3;

QY 59 EAGRIRGPRVGNELCEVHYLDLCRIRESYSE-LQLEAKKQELNSELAKLNKLTIEAC 117
 Db 43 EREDDLEKLAQAQCSH---LMREHEDVQERTTLRYEBRITELSHVIAELNKKIDRL 99

QY 118 KKSIEAKODLQKLVISQTEHSYKE---LMAONQPKLSLP 156
 Db 100 QGTTIREDEYSELRSLSOSQHEVNEDESRMDDQTSVSLP 141

US-08-446-549-2
 ; Sequence 2, Application US/08446549
 ; Patent No. 5693536

GENERAL INFORMATION:
 APPLICANT: Vogelstein, Bert
 APPLICANT: Kinzler, Kenneth W.
 APPLICANT: White, Raymond
 APPLICANT: Nakamura, Yusuke
 TITLE OF INVENTION: Gene Mutated in Colorectal Cancer of
 TITLE OF INVENTION: Humans
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner, Birch et al.
 STREET: 1001 G Street
 CITY: Washington
 STATE: D.C.
 COUNTRY: U.S.A.
 ZIP: 20001-4597

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/446,549
 FILING DATE: 19-MAY-1995
 CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/670,611
 FILING DATE: 13-MAR-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.33981

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 829 amino acids
 TYPE: amino acid
 STRANDEDNESS: single

TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ANTI-SENSE: NO

ORIGINAL SOURCE:
 ORGANISM: Homo sapiens

US-08-446-549-2

Query Match 2.1%; Score 102.5; DB 1; Length 829;
 Best Local Similarity 29.4%; Pred. No. 0.55;
 Matches 30; Conservative 22; Mismatches 43; Indels 7; Gaps 3;

QY 59 EAGRIRGPRVGNELCEVHYLDLCRIRESYSE-LQLEAKKQELNSELAKLNKLTIEAC 117
 Db 43 EREDDLEKLAQAQCSH---LMREHEDVQERTTLRYEBRITELSHVIAELNKKIDRL 99

QY 118 KKSIEAKODLQKLVISQTEHSYKE---LMAONQPKLSLP 156
 Db 100 QGTTIREDEYSELRSLSOSQHEVNEDESRMDDQTSVSLP 141

Search completed: September 4, 2002, 14:57:37
 Job time: 37 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 4, 2002, 14:57:00 ; Search time 35.77 Seconds
(without alignments)
2853.699 Million cell updates/sec

Title: US-09-809-920-4
Perfect score: 4873
Sequence: 1 MTGYTMLRNGAGNGGQTCM.....DSVLEFTRLPHDKTKCFKFI 919

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: A_Geneseq_032802.*
2: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1999.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT.*
23: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4873	100.0	919	20	AAV08481 Human exostosis-4
2	4873	100.0	919	21	AAV84669 Amino acid sequence
3	4755	97.6	919	22	AAV84668 Rat Reg-binding pr
4	4720.5	96.9	918	21	AAV84668 Amino acid sequence
5	2314	47.5	972	22	ABV70158 Drosophila melanog
6	1666.5	34.2	364	22	AAV84675 Rat Reg-binding pr
7	1436	29.5	262	21	AAV84675 Amino acid sequence
8	1000	20.5	718	20	AAV28883 Human Ext-2 protei
9	998	20.5	663	21	AAV85946 Breast and ovarian
10	939	19.3	728	19	AAV44851 Hereditary multipl
11	871	17.9	746	20	AAV28862 Human Ext-1 protei

12	848.5	17.4	259	21	AAV84680 Amino acid sequence
13	841	17.3	717	22	ABB61749 Drosophila melanog
14	796.5	16.3	760	20	AAV28884 Drosophila Tcy pro
15	795.5	16.3	760	22	ABB64881 Drosophila melanog
16	663.5	13.6	269	21	AAV84676 Amino acid sequence
17	625	12.8	125	21	AAV84670 Amino acid sequence
18	530	10.9	270	21	AAV84677 Amino acid sequence
19	399.5	8.2	262	21	AAV84678 Amino acid sequence
20	367	7.5	330	21	AAV82363 Human alpha-1,4-N
21	365	7.5	270	21	AAV84679 Amino acid sequence
22	290	6.0	329	21	AAV84673 Arabidopsis thalia
23	199.5	4.1	123	21	AAV84673 Amino acid sequence
24	197.5	4.1	124	21	AAV84672 Amino acid sequence
25	180	3.7	121	21	AAV84674 Amino acid sequence
26	176.5	3.6	120	21	AAV84671 Amino acid sequence
27	160	3.3	103	20	AAV60148 Human endometrium
28	142	2.9	161	22	AAV01937 Human polypeptide
29	141	2.9	181	22	ABG30294 Novel human diapo
30	123	2.5	1378	22	ABG23678 Novel human diapo
31	123	2.5	1400	22	ABG09151 Novel human diapo
32	123	2.5	1788	22	ABG06749 Novel human diapo
33	123	2.5	1921	22	ABG06576 Novel human diapo
34	123	2.5	2563	22	ABG14767 Novel human diapo
35	122.5	2.5	537	21	AAV14498 Arabidopsis thalia
36	122.5	2.5	549	21	AAV14497 Arabidopsis thalia
37	121.5	2.5	415	21	AAV07301 Arabidopsis thalia
38	121.5	2.5	415	21	AAV46602 Arabidopsis thalia
39	121.5	2.5	423	21	AAV46602 Arabidopsis thalia
40	121.5	2.5	423	21	AAV41326 Arabidopsis thalia
41	121.5	2.5	755	22	ABG04875 Novel human diapo
42	120.5	2.5	2633	22	ABG06505 Novel human diapo
43	119.5	2.5	2901	22	ABG09163 Novel human diapo
44	119	2.4	613	22	ABG06752 Arabidopsis thalia
45	118.5	2.4	412	21	AAV26205 Arabidopsis thalia

ALIGNMENTS

RESULT 1	
AAV08481	
ID	AAV08481 standard; Protein; 919 AA.
AC	AAV08481;
DT	29-JUL-1999 (first entry)
XX	
DE	Human exostosis-4 protein.
XX	
KW	Exostosis-4; human; cytostatic; immunosuppressive; anti-HIV; Ext-2;
KW	treatment; cancer; tumour; genetic disorder; AIDS; diagnostic; screening;
KW	acquired immune deficiency syndrome.
XX	
OS	Homo sapiens.
XX	
PN	WO925822-A1.
XX	
PD	27-MAY-1999.
XX	
PF	14-NOV-1997; 97MO-CN00126.
XX	
PR	14-NOV-1997; 97MO-CN00126.
XX	
PA	(UYHU-) UNIV HUNAN MEDICAL.
XX	
PI	Deng H, Fan C, Liu C, Ruan Q, Xia J, Xu L;
XX	
DR	WPI: 1999-347470/29.
DR	N-PSDB: AAV72378.
XX	
PT	New Exostosis-4 gene useful in the treatment of cancer, tumors,
XX	genetic disorders and AIDS

PS Claim 1; Page 24-27; 33pp; English.

CC This invention describes a novel human Exostosin-4 polypeptide which has
 CC cytosolic, immunosuppressive and anti-HIV activity. Exostosin-4 has
 CC homology with Ext-2 and is therefore expected to have similar biological
 CC activities. Exostosin-4 and its modulators can be used for treatment of
 CC cancer, tumours, genetic disorders and AIDS (Acquired Immune Deficiency
 CC Syndrome), and in diagnostic/screening assays for such conditions.

XX Sequence 919 AA:

Query Match 100.0%; Score 4873; DB 20; Length 919;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 919; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGYMLRNGAGNGGQCMLRMSNRIRLTWLSFTLVLFVPEPLIAHYLTITLDEADA 60
 DB 1 mgytmlrngagnggqctmlrwnrirltwlsftlvlfvpepliahyltltldeada 60
 QY 61 GKRIGPRVGNELCEVKHVLDCRIRSVSEELLQLEAKROELNSEIAKLNKIEACKS 120
 DB 61 gkrigprvgnelcevkhlvdlcrirsvseellqlleakroelnseiaklnkieacks 120
 QY 121 IENAKQDLLQKLVISQTEHSHYKELMAQNOPKLSPIRLLPEKDDAGLPPEKATGCRH 180
 DB 121 ienakqdllqlknvisqtehsykelmaqnpklsipirllpekddaglppekattrgcrh 180
 QY 181 NCFDYSRCPLTSGFPVYVYDSQPFVGSYLDPVKQAFQATARANVYTENADIACLYVI 240
 DB 181 ncfdysrcpltsfgfpvyvdsqpfvgsyldpvlkqafqataranvytenadiaclyvi 240
 QY 241 LVGEKQEPVLRPALEKOLYSLPHWRDGNHVTINLSRSDPNQLXNSTGAMVAQ 300
 DB 241 lvgekepvvlrpaalekolslphwrtdgnhvtinlsrdsdpnqlxnstgamaaq 300
 QY 301 STFTYVOYRPGFDLVSPLVHAMSEPNFMEIPQVPVARKYLEFTQGEKISLSLSQEA 360
 DB 301 stftvyoyrpgfdlvspplvhamsenpfmepqvpyarkylftqgekislslsqea 360
 QY 361 RSFEEMBGDPADYDRIITLKAVDSDKLDQVLFVETCKNOKPSLPTFMALCGERED 420
 DB 361 rsfeembgdppadydrilrtlkavdsdkldqylfvetcknkpkslptfmalcgered 420
 QY 421 RLLELKSTFALITTPGPRLVISGCASTRLEALEVGAVVYVGEQVLPYQMLQNE 480
 DB 421 rllelkstfaliitpgdprlvissgcatrlealevgavvyvgeqvlpyqmlqne 480
 QY 481 AALVVPKRVTEVHFLRLSDSDLLAMRQGRFLMETSTADISFNVTYLAIRTRIOI 540
 DB 481 aalvvpkrvtevhflrlsdsdllaamrqgrflmetstadsifnvtylamirtrioi 540
 QY 541 PAAPIRREAAAEIRPHRSGKAAGTPPNADNGDDLGVETETPPASPRYLNFETLVTDF 600
 DB 541 paapirreaaaeirphrsgkaagtppnadngddlgveteppasprylnftlvtdf 600
 QY 601 YRSWNCAPGCFHLEPFRPPVLPSEAKELSGSGRPRDGGAGSGSKERQALGVNPR 660
 DB 601 yrswncapegfhlepfrppvlpseakelsgsgprpddggagsgskerkqalgvnpr 660
 QY 661 EQFTVVMVLYERBEVLNLSLERLNGLPYLNKVVVVMNSPKLPSDDLMPDIGNVIVRT 720
 DB 661 eqftvvmvlyerebevlnlslerlnglpylnkvvvvnspklpsddlmpdignvivr 720
 QY 721 ERKSLNNRFLPMEIETEAALISIDDAHLRHDEIMFGFRVREARDRIVGFGRYHAMDI 780
 DB 721 erkslnnrflpmeieteaalisiddahlrhdeimfgfrvreardrvfgfryhamdi 780
 QY 781 PHOSWLYNSVSGELSMVLTGAFFHKYAYLYSYVMPQAIRMVDEYINCEDIAMFLV 840
 DB 781 phoswlynsvglsmvltgaaffhkyaylysyvmpqairmvdeyincediameflv 840
 QY 841 SHITRKPIKVTSRMTFRCPGCPQALSHDSDSHFHERHKCNINFFVKKYGYMPLIYTGFRVD 900

DB 841 shitrkpiikvtsrmtfrpcpqlshdsdshfherhcninffvkkygympliytgfrvd 900
 QY 901 SVLEKTRLPHPDKTCRKF 919
 DB 901 svlcktrlphtkcktkf 919

RESULT 2

AA84669
 ID AAY84669 standard; Protein; 919 AA.

AC AAY84669;

DT 25-JUL-2000 (first entry)

DE Amino acid sequence of a human TRRX protein.

DE Tumour necrosis factor receptor-associated Factor; TRAF;
 KW TRF-protein-interacting hereditary multiple extoses protein; TRRX;
 KW signal modulator; tumour necrosis factor receptor;
 KW CD40 mediated signal transduction; TRAF protein; cancer;
 KW hereditary multiple extosis; autoimmune disease.

XX Homo sapiens.

XX WO200018959-A1.

XX 06-APR-2000.

XX 17-SEP-1999; 99WO-US21654.

XX 17-SEP-1998; 98US-0156191.

XX (VXCO) UNIV COLUMBIA NEW YORK.

XX Sato T;

XX WPI; 2000-293180/25.

XX N-PSDB; AAA12734.

XX New nucleic acid encoding Tumour necrosis factor Receptor-Associated
 PT Factor (TRAF) protein-interacting hereditary multiple extoses (TRRX)
 PT protein, useful in the diagnosing cancer -

XX Claim 39; Fig 8A-B; 161pp; English.

XX The present sequence represents a human tumour necrosis factor receptor-
 CC associated factor (TRAF) protein-interacting hereditary multiple
 CC extoses (TRRX) protein. TRRX is a signal modulator which bridges
 CC between the tumour necrosis factor (TNF) receptor and CD40 mediated
 CC signal transduction. Anti-TRRX antibodies are useful for treating an
 CC abnormality in a patient by inhibiting binding of a TRRX protein and
 CC a TRAF protein (that is, TRAF 2, TRAF 3 or TRAF 5). The abnormality
 CC is cancer, a hereditary multiple extosis or an autoimmune disease. The
 CC cancer is colon cancer, gastric cancer, human head and neck squamous
 CC cell carcinoma, prostate carcinoma, breast cancer, thyroid cancer,
 CC oesophageal cancer, lung cancer, colorectal cancer, ovarian cancer,
 CC papillary bladder cancer, osteosarcoma, chondrosarcoma, liposarcoma,
 CC giant cell tumour, Ewing sarcoma, or other malignant tumours.

XX Sequence 919 AA:

Query Match 100.0%; Score 4873; DB 21; Length 919;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 919; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTGYMLRNGAGNGGQCMLRMSNRIRLTWLSFTLVLFVPEPLIAHYLTITLDEADA 60
 DB 1 mgytmlrngagnggqctmlrwnrirltwlsftlvlfvpepliahyltltldeada 60
 QY 61 GKRIGPRVGNELCEVKHVLDCRIRSVSEELLQLEAKROELNSEIAKLNKIEACKS 120

Db 61 gkrlfgrvgnelcevhvhdicrlresvseellqleakrqelnsejaklnlkeacks 120
121 IENAKODLLOLKNVISTEHSYKELMAQONPKLSLPIRLPEKDDAGLPPPKATRCGRlh 180
121 IENAKODLLOLKNVISTEHSYKELMAQONPKLSLPIRLPEKDDAGLPPPKATRCGRlh 180
121 IENAKODLLOLKNVISTEHSYKELMAQONPKLSLPIRLPEKDDAGLPPPKATRCGRlh 180
181 NCFDYSRCPLTSGFPVYVSDQFVFGSYLDPLVKQAFQATARANYVTENADIACLYVI 240
181 nctdyrcpltsfgfprvyvsdqfvgfsgyldplvkqafqataranyvtenedaclyvi 240
241 LVGEMOPVVLRAPELEKOLYSLPHMRTDGNHNTILNSKSDQNLVNVSTGRANVAQ 300
241 lvgemopvvlrapelekolyslphmrtdgnhntilnsksdqnlvvnstgranvaq 300
301 STEFTVQYRRGFDLVSPVLYHAMSEPNFMETPPQVPRKRYLFTFGCEKTESLRSIQEA 360
301 stftvqyrrgfdlvspvlyhamsepnfmetsppqvprkrylftfgcektesslrsiqea 360
361 RSFEEMEGDPPADYDRIATATLKAQVDSKLDQVLEFTCKNPKPSLPTEMALCGERD 420
361 rsfeemegppadydrilatatlkaqvdskldqvleftcknppkslptemalcgere 420
421 RLELLKSTFALITTPGDPRLVSSGCATRLFEALEVGAAPVVLGEVOLPYODMLQNE 480
421 rlellkstfaliitpgdprlvssgcattrlfealevgavpvvlgevolpyodmlqne 480
481 AALVPRKRVTEVHFLRLSLSDSLAMRQGRFLMETYESTADSIPTNYLAMIRRIQI 540
481 aalvprkrtvtehflrlslsdslamrqgrflmetyestadsipntylamirriqi 540
541 PAARIREEAAAEIHRSGKAAGTDPNMADNGDLDGVEYEPPIASRYLRNFTLYTDF 600
541 paarireeaaaeihrsgkaagt d pnmadngdldgveypipiasrylrnftlytdf 600
601 YRSNCAKAPGPHLPHPFPFVLPSEAKFLSGNGFPIGAGSGSKEFOALGAVNPR 660
601 yrsncakapghlphfpfpvlpseakflsgngfpiagsgskefoalgvnpr 660
661 EQFTVMALYERESEVNLNSLERLNGRLPYLNKVVVWNSPKLPSBDLMPDIGVIMVVRT 720
661 eqftvmalyeresevnlnslerlngrlpylnkvvnwvnsplpsedlmpdigvimvrt 720
721 EKSNLNRPLPMNIEFETALISIDDAHLKHDIMEGFRVWRKADRIVGPFGYHAMDI 780
721 eksnlnrplpmniefetalisiddahlkhdimegfrvwrkadrivgpfgyhamdi 780
781 PHOSMLNSVSCELSMVLGAAPFHKYVAYLYSVMPQALRDMVDYINCEDIAMNELV 840
781 phosmlnsvscelsmvlgaapfhkyvaylysvmpqalrmdvdyincediameelv 840
841 SHITRRPPIKVTSMWTRCPCPOLASHDSHFHERHKCNFVKKVGYMPLLYTQFRVD 900
841 shitrppikvtsmwtrcpcpolashdshfherhkcnfvkkygmpllytqfrvd 900
901 SVLFKTRLPKDKTCKEFTI 919
901 svlfktrlpkdktkcfti 919

RESULT 3
AAB51341 standard: Protein: 919 AA.
XX
XX AAB51341:
XX
XX 05-APR-2001 (first entry)
XX
XX Rat Reg-binding protein SEQ ID NO:4.
XX
XX Rat: Reg-binding protein; Reg receptor; Reg: DNA synthesis; apoptosis;
XX
XX cell proliferation; diabetes; antidiabetic.

OS Rattus norvegicus.
PN WO200077192-A1.
XX
XX 21-DEC-2000.
XX
XX 09-JUN-2000; 2000WO-JP03764.
XX
XX 10-JUN-1999; 99JP-0164488.
XX
XX (OKAM/) OKAMOTO H.
PI Okamoto H;
XX
XX WPI: 2001-061871/07.
DR N-PSDB; AAF26844.
XX
PT Protein binding to pancreatic Reg protein and its encoding DNA, useful
XX for screening candidate antidiabetic drugs
XX
PS Claim 1; Page 74-81; 11pp; Japanese.
XX
XX The present invention describes DNA which encodes: (a) a protein of rat
XX origin binding to Reg protein; or (b) a Reg-binding protein derived from
XX the natural Reg sequence by addition, deletion and/or substitution of
XX one or more amino acid residues. The present sequence represents a rat
XX Reg-binding protein from the present invention. Reg-binding proteins
XX have antidiabetic activity. Drug compositions comprising Reg-binding
XX proteins and their genes are useful for the treatment and/or prevention
XX of diabetes. In addition, they are useful in developing new treatments
XX for diabetes.
SQ Sequence 919 AA:

Query Match 97.6%; Score 4755; DB 22; Length 919;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 894; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY 1 MTGYTLNRNGAGNGGOTCMLRMSNRIRLTMSTFLVILVFPPLAHVYLTLDDEADEA 60
1 mtgytlrnrgagnggotcm lrmsnrir ltmstflv ilvfpplahvy ltldeadea 60
Db 1 mtgytlrnrgagngggtcm lrmsnrir ltwlftlilvfpplahy ltldeadea 60
QY 61 GKRIFGPRVGNELCEYKHNVDLCRIRESVSEBELQLEAKQONLSETAKNLKTEACKS 120
61 gkrlfgrvgnelcevhvhdicrlresvseellqleakrqelnsejaklnlkeacks 120
Db 61 gkrlfgrvgnelcevhvhdicrlresvseellqleakrqelnsejaklnlkeacks 120
QY 121 IENAKODLLOLKNVISTEHSYKELMAQONPKLSLPIRLPEKDDAGLPPPKATRCGRlh 180
121 IENAKODLLOLKNVISTEHSYKELMAQONPKLSLPIRLPEKDDAGLPPPKATRCGRlh 180
121 IENAKODLLOLKNVISTEHSYKELMAQONPKLSLPIRLPEKDDAGLPPPKATRCGRlh 180
181 NCFDYSRCPLTSGFPVYVSDQFVFGSYLDPLVKQAFQATARANYVTENADIACLYVI 240
181 nctdyrcpltsfgfprvyvsdqfvgfsgyldplvkqafqataranyvtenedaclyvi 240
Db 181 nctdyrcpltsfgfprvyvsdqfvgfsgyldplvkqafqataranyvtenedaclyvi 240
QY 241 LVGEMOPVVLRAPELEKOLYSLPHMRTDGNHNTILNSKSDQNLVNVSTGRANVAQ 300
241 lvgemopvvlrapelekolyslphmrt d gn hntilnsksdqnlvvnstgranvaq 300
Db 241 lvgeigepavllqpadlekqlhslphwrt d gnhnvlilnsksdqnllyvstgramaq 300
QY 301 STEFTVQYRRGFDLVSPVLYHAMSEPNFMETPPQVPRKRYLFTFGCEKTESLRSIQEA 360
301 stftvqyrrgfdlvspvlyhamsepnfmetsppqvprkrylftfgcektesslrsiqea 360
Db 301 stftvqyrrgfdlvspvlyhamsepnfmetsppqvprkrylftfgcektesslrsiqea 360
QY 361 RSFEEMEGDPPADYDRIATATLKAQVDSKLDQVLEFTCKNPKPSLPTEMALCGERD 420
361 rsfeemegppadydrilatatlkaqvdskldqvleftcknppkslptemalcgere 420
Db 361 rsfeemegppadydrilatatlkaqvdskldqvleftcknppkslptemalcgere 420
QY 421 RLELLKSTFALITTPGDPRLVSSGCATRLFEALEVGAAPVVLGEVOLPYODMLQNE 480
421 rlellkstfaliitpgdprlvssgcattrlfealevgavpvvlgevolpyodmlqne 480
Db 421 rlellkstfaliitpgdprlvssgcattrlfealevgavpvvlgevolpyodmlqne 480
QY 481 AALVPRKRVTEVHFLRLSLSDSLAMRQGRFLMETYESTADSIPTNYLAMIRRIQI 540
481 aalvprkrtvtehflrlslsdslamrqgrflmetyestadsipntylamirriqi 540
XX

Db 481 aalvvpkprvtevhflrlsdsdillamrgrflwetyfstadsifntvlamirtrig 540
 QY 541 PAAPIREEAAAEIPHRSGKAAGTDPNADNGDLGPEVTEPPYASPRYLNNFTLVTFD 600
 Db 541 paapireeaaaeiphrsgkaagtdpnmadngdlgpevteppypasprylntflvtldc 600
 QY 601 YRSMNCAPGFHLFPHPTFPDPLPSEAKFLGSGTGFPRIGGAGSGKEFOALGNNPR 660
 Db 601 yrwmsapgfhlfnhpcfpdvlpseakflgsqgtftrpggagsgskelgaa19gnvqr 660
 QY 661 EGFVVMVLTREBEVLANSLERLNGLPYLKVVVWVNSPKLPSEDLMPDGVPIWVPT 720
 Db 661 egftvmltyereevlmslerlnglpylnkvvwvnsmpklpsedllmpdgvpiwvrt 720
 QY 721 EKNLSNNRFLPWNIEETALISIDDAHLRHEIMFGFRWREARDRIVGFGRYHAMD 780
 Db 721 eknslnnrflpwnieteaalisiddahlrndelmfgfrwreardrivgfgyrhwadl 780
 QY 781 PHOSWLYNSNSCELSMWLTGAFFHKYAYLYSYVMPQATRDVDEYINCEDIAMNELY 840
 Db 781 phgswlynsnyscelsmwltgaafhkyyaylysyvmpqatrdmveyincedlamnflv 840
 QY 841 SHITRRPDIKYSRMTFRCPCGPQALSHDSEHFERHAKINEFVYVGYMPLTYQFVND 900
 Db 841 shitrkppikvtsrwtfcgpcgpqalshdshferhkcinfvkvgymplytqftrvd 900
 QY 901 SVLRKTRLPHKTKCFKI 919
 Db 901 svlkrtrlphtkctkcfki 919

RESULT 4

AA84668
 ID AA84668 standard; Protein: 918 AA.

AA84668:

25-JUL-2000 (first entry)

Amino acid sequence of murine TRFX protein.

XX Tumour necrosis factor receptor-associated Factor; TRAF;
 KW TRF-protein-interacting hereditary multiple extoses protein; TREX;
 KW signal modulator; tumour necrosis factor receptor;
 KW CD40 mediated signal transduction; TRAF protein; cancer;
 KW hereditary multiple extosis; autoimmune disease.

XX Mus sp.

OS WC200018959-A1.

PN 06-APR-2000.

XX 17-SEP-1999; 99WO-US21654.

XX 17-SEP-1998; 98US-0156191.

XX (UYCO) UNIV COLUMBIA NEW YORK.

XX Sato T;

PI WPI; 2000-293180/25.

XX N-PSDB; AAA12733.

XX New nucleic acid encoding Tumour necrosis factor Receptor-Associated
 PT Factor (TRAF) protein-interacting hereditary multiple extoses (TREX)
 PT protein, useful in the diagnosing cancer -

XX Claim 38; Fig 7A-B; 161pp; English.

XX The present sequence represents a murine tumour necrosis factor receptor-
 CC associated factor (TRAF) protein-interacting hereditary multiple
 CC extoses (TREX) protein. TREX is a signal modulator which bridges

CC between the tumour necrosis factor (TNF) receptor and CD40 mediated
 CC signal transduction. Anti-TREX antibodies are useful for treating an
 CC abnormality in a patient by inhibiting binding of a TREX protein and
 CC a TRAF protein (that is, TRAF 2, TRAF 3 or TRAF 5). The abnormality
 CC is cancer, a hereditary multiple extosis or an autoimmune disease. The
 CC cancer is colon cancer, gastric cancer, human head and neck squamous
 CC cell carcinoma, prostate carcinoma, breast cancer, thyroid cancer,
 CC oesophageal cancer, lung cancer, colorectal cancer, ovarian cancer,
 CC papillary bladder cancer, osteosarcoma, chondrosarcoma, liposarcoma,
 CC giant cell tumour, Ewing sarcoma, or other malignant tumours.
 CC XX
 SQ Sequence 918 AA:

Query Match 96.9%; Score 4720.5; DB 21; Length 918;
 Best Local Similarity 97.0%; Pred. No. 0;
 Matches 891; Conservative 7; Mismatches 20; Indels 1; Gaps 1;

QY 1 MTGYMLRNGAGNGGOTCMRWSNRRLTWLSFTLVILVFPPLIAHYLTIDEADEA 60
 Db 1 mtygmrlngvgngggqcmrwsnrirltwlsftllvffflahytltddeadea 60
 QY 61 GKRTFGPVGNGELCEVKHVLDCRIRSVSEELLQLEAKRQELMSEIATKMLTEACKKS 120
 Db 61 gkrltfgpagselcevkhnvlcrlresvseellqlleakrqlnsetaklnlkteacks 120
 QY 121 IENAKODLLQKLVNYSQTEHSYKELMAONQPKLSPIRLPEKDDAGLPPKATRGCLH 180
 Db 121 ienakodllqklnvysqtehsykelnmgngklsiprllpckddaglppektrgclh 180
 QY 181 NCFDYSRCPLTSGFPVYVYDSQDFEGSYLDPVKQAFOATARANVYVENADIACLYVI 240
 Db 181 ncfdyrcpltsghfpvyvysdqdfegsyldpvlkqafqatranvyytenadiaclyvv 240
 QY 241 LVGMEQEPVYLRPALEKOLYSLPHWRDGNHNIINLSRSQTNILYNSSTRANVAQ 300
 Db 241 lvgmeqepvylrpadlekqlslphwrdgnhnhlilstsdqtnillynstranvagr 299
 QY 301 STEFYVOYRPGFDLVSPVHAMSEPNMETPPQVPRKRYLFFQGEKTESLRSSLOEA 360
 Db 300 stfyaqyragfdlvspvhamsepnmetppqvpkrkyllffqgekestlrsslgea 359
 QY 361 RSFEENBGGPPADYDRIATLKAVDSKLDQVLEFTCKNDKPPSLPTHEMALCGERED 420
 Db 360 rsfeenbgppadydrilatlkavqskldqylveftcknqkpslpwemalcgred 419
 QY 421 RLELKLSTFALITTPGDPRLVYSSGCATRLEALEAGAVPVYLGEOVQPYQDMQWNE 480
 Db 420 rlellkfstallittpgdprrllssgcatrilealegavpvylgevgqlpyndmlqwe 479
 QY 481 AALVVPKRVTEVHFLLRSLSDSDDLAMRRQGRFLMETVSTADSIENTVIAMIRFIQI 540
 Db 480 aalvvpkprvtevhflrlsdsdillamrrqgrflwetyfstadsifntvlamirtrig 539
 QY 541 PAAPIREEAAAEIPHRSGKAAGTDPNADNGDLGPEVTEPPYASPRYLNNFTLVTFD 600
 Db 540 paapireeaaaeiphrsgkaagtdpnmadngdlgpevteppypasprylntflvtldc 599
 QY 601 YRSMNCAPGFHLFPHPTFPDPLPSEAKFLGSGTGFPRIGGAGSGKEFOALGNNPR 660
 Db 600 yrwmsapgfhlfnhpcfpdvlpseakflgsqgtftrpggagsgskelgaa19gnvqr 659
 QY 661 EGFVVMVLTREBEVLANSLERLNGLPYLKVVVWVNSPKLPSEDLMPDGVPIWVPT 720
 Db 660 egftvmltyereevlmslerlnglpylnkvvwvnsmpklpsedllmpdgvpiwvrt 719
 QY 721 EKNLSNNRFLPWNIEETALISIDDAHLRHEIMFGFRWREARDRIVGFGRYHAMD 780
 Db 720 eknslnnrflpwnieteaalisiddahlrndelmfgfrwreardrivgfgyrhwadl 779
 QY 781 PHOSWLYNSNSCELSMWLTGAFFHKYAYLYSYVMPQATRDVDEYINCEDIAMNELY 840
 Db 780 phgswlynsnyscelsmwltgaafhkyyaylysyvmpqatrdmveyincedlamnflv 839

Oy		84.1 SHIRKPRIVXTSMTEFRCPCSPQALSHDDSHFERHRRHCINFRFAVYVCMTLLTYQGFVD	900
Dg		840 shlrkprkvksrvtfcpcgpelshddsnfherhkcinffvkvygmpllytclqfrvd	899
Oy		901 SVLEFKRLPHDKTKCFKKFI	919
Dg		900 svlfkrlphdktkcfkffi	918
RESULT	5		
ID	ABB70158		
XX	ABB70158 standard; Protein; 972 AA.		
AC	ABB70158;		
XX			
DT	26-MAR-2002 (first entry)		
XX			
XX	Drosophila melanogaster polypeptide SEQ ID NO 37266.		
XX			
XX	Drosophila: developmental biology; cell signalling; insecticide;		
KW	pharmaceutical.		
XX			
OS	Drosophila melanogaster.		
XX			
PN	MO200171042-A2.		
PD			
XX	27-SEP-2001.		
PF			
XX	23-MAR-2001; 2001WO-USO9231.		
PR			
XX	23-MAR-2000; 2000US-191637P.		
PA			
PI	11-JUL-2000; 2000US-0614150.		
PJ	(PEKE) PE CORP NY.		
PT	Venter JC, Adams M, Li PWD, Myers EW;		
PS	WPI: 2001-6556860/75.		
XX	N-PSDB: ABLI4261.		
XX			
XX	New isolated nucleic acid detection reagent for detecting 1000 or more		
XX	genes from Drosophila and for elucidating cell signalling and cell-cell		
XX	interactions -		
XX	Disclosure: SEQ ID NO 37266; 21pp + Sequence Listing; English.		
XX			
CC	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention is		
CC	useful in developmental biology and in elucidating cell signalling and		
CC	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA		
CC	sequences (ABB57737-ABB72072).		
CC	(ABB57737-ABB72072).		
CC	The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
CC			
XX	Sequence 972 AA;		
SQ			
Query Match	47.5%: Score 2114; DB 22: Length 972;		
Best Local Similarity	48.1%: Pred. No 9.5e-211;		
Matches 473:	Conservative 153; Mismatches 252; Indels 106; Gaps 21;		
Oy		11 GAGNGGQC-----MLRWSNRRLTW-----LSFTLEVLIVFPPLAHVYLTLTD	55
Dg		11 :	
		20 gsggnacapssaaqlrthsngftswmrgtrykklpmvllmlflvsclayillsveq	79
Oy		56 EAD-----BAGKRIFGRVGNELCEVKHHVLDLCRIRESSESLQLQEARQELN	104
Dg		80 dappldlnrseplldaysdfsaamrgdlkmrie--emwrlkstsvsvelrelstsrqkfq	136

OY	105	SEIKALNLKTEACKKSLSENNAKODLDLKNVISTOHSYKELMAQONPKSLPTRIILLPEKD	164
Db	137	sdlsqgnqkileekqellleqtelterlklsveqayqaeayvnrnfpdlaiprslpntc	195
OY	165	DAGLPPPK-----ATGCRCLAHNCFDYSRCLPSGPPVYVSDQGVF---GSYDLPLVK	215
Db	196	---lprksnpltgmaascemhncfnhsrctslsgfpvylpdpdshsvgrkgydlldgflk	252
OY	216	QAFQATARANYVTENADLCYLVIWGE-MQBPVYL-----	252
Db	253	tlkqetlgyhahlvkdpkhacilylvlgaeallgeqdlrlnrnyaaqaehqgsrptclend	312
OY	253	-PAELEKQLSLSPMKRWDGNHNYILNLSK---SDQLNLNVSTGRAMVAAOSTFYTOY	308
Db	313	cpvmek-lyslpwyggdgqnhvhlmlarrdltshtlmplyrqntnlralvvasafereqf	371
OY	309	RGEFDLVSPVLVHMASDPNMEIRPOVPVRYKTLFTEOGKIESLSSSQEARSFEEME	368
Db	372	rpygdlllvprllpprggdvvqceamyparkyllvygqe-lrpkysslnpidafileh	430
OY	369	GDPADYDDRIATLKAANODSLDOVLVEFTC---KNPKRSLPTEMALGSEREDREL	424
Db	431	ad-----makgatqdfyqlfgcvateqegedslp-dwlcgsdsarsql	475
OY	425	LKLSTFALITTRGDPRLVYISSGCASTRLEALVEYGAAPVVLG-EQVQLPRLQMLQHNQAL	483
Db	476	lkdfstslilprlprng-rvsstlmairyeelrsgavpvlvgadelrlpiaevedvrtel	534
OY	484	VVPPRVATVEVFLRLSDSDLDLAMRHOGAFLEWETSTADSIFFNVLMITRRIQIPAA	543
Db	555	llparitelhflrravgdadlllrrtqglllwerlylssvqavdclvslsldrlgiprr	594
OY	544	PIREBAAE-----IHRSGKAAGTDPNNADNGDLDLGVTETEPYPASPRYLRFNFTLV	597
Db	595	pypsvlqsvfnstflrlkxdprrygdtepees----lgrl--eppypapafrrnylllr	648
OY	598	TDFRSMWCAAGRHRLPHNPRPDVRLPSEAKFLGSGTGRPPGAGGSGSGKEQMAQGN	657
Db	649	mqakeawndwldrplylvpqrlpdpalprseakfmgshntgfrlp9kxlg9gakefgeslgn	708
OY	658	VPRQETVVMVLTREBVLNLSLERLNGTLRYLKNVVVWNSPKLPSEDLMPDIGVPIW	717
Db	709	ypreqfctvmlytqereglmdslsgrllygrlyllkvvvvwnspkrpdlrltwpdlgyprav	768
OY	718	VRTKNSLNNRFLPWNDETFEALISIDDOANLKHDEIMEGFRVWRAPARDRIVGPGRTYHA	777
Db	769	lrapsnslnflrfdyaleaeavlsvddahllhdelllfgfrwrehtgrvvgfprgyna	828
OY	778	WDI---PHOSMLYNSYNSACELSMVLTGGAARFKYUATLYLVAMPQALRDWAVDEYICEDIA	835
Db	829	wdlgnpngqwhynyscelsmwlvgaaivkhykyllylvtrhlpqalrtldvdeimmedla	888
OY	836	MNPLVSHITRKPRIKVYSRMTFRQPGCPQALSHDSDHFHERHKCNIFEVKKVYGMPLLYT	895
Db	889	mnflvshlitrkprkvksrvtctfcpgspvaslseddthfgerlnkcnlffarfygycrplnt	948
OY	896	QFRVDSVLFKTRLPHDXTKCFKFL	919
Db	949	qyradsllfkrtrlpdkqkcfkyl	972
RESULT 6			
ID	AA051340	standard; Protein: 364 AA.	
XX	AA051340:		
XX	AC		
XX	DT	05-APR-2001 (first entry)	
XX	Rat	Reg-binding protein SEQ ID NO:2.	
XX	Rat	Reg-binding protein SEQ ID NO:2.	
XX	Rat	Reg-binding protein; Reg receptor; Reg; DNA synthesis; apoptosis;	

KW cell proliferation; diabetes; antidiabetic.
XX
XX Rattus norvegicus.
OS
XX WO200077192-A1.
XX
XX PD 21-DEC-2000.
XX
XX 09-JUN-2000; 2000WO-JP03764.
PF
XX 10-JUN-1999; 99JP-0164488.
XX
XX (OKAM.) OKAMOTO H.
XX
XX OKamoto H;
PI
XX WPI: 2001-061871/07.
XX
XX N-PSDB; AAF26843.
DR
XX
XX Protein binding to pancreatic Reg protein and its encoding DNA, useful
PT for screening candidate antidiabetic drugs -
PS
XX Claim 1; Page 61-64; 11pp; Japanese.
XX
XX The present invention describes DNA which encodes: (a) a protein of rat
CC origin binding to Reg protein; or (b) a Reg-binding protein derived from
CC the natural Reg sequence by addition, deletion and/or substitution of
CC one or more amino acid residues. The present sequence represents a rat
CC Reg-binding protein from the present invention. Reg-binding proteins
CC have antidiabetic activity. Drug compositions comprising Reg-binding
CC proteins and their genes are useful for the treatment and/or prevention
CC of diabetes. In addition, they are useful in developing new treatments
CC for diabetes.
XX
XX Sequence 364 AA;
SQ
Query Match 34.2%; Score 1666.5; DB 22; Length 364;
Best Local Similarity 85.4%; Pred. No. 9.8e-150;
Matches 323; Conservative 9; Mismatches 19; Indels 27; Gaps 2;
XX
1 MTGTYMLRNGAGNGGOTCMLRMSNRIRLMLSPFLVILVFPPLIAHYITLDEADEA 60
Db 1 mtygtmlrngvgvngqctcmrlwsmrirltwlslfllvfppliahyytlcdeaada 60
XX
61 GKRIFFRGVNELCENVHLDLCRIRSVSEELLQLEAKQELNSETAKNLKTEACKS 120
Db 61 gkrlffrgvnelcevhldlcrirsvseellqleakqelnselaknlkteacks 120
XX
121 IENAKODLQKLVNISTQTEHSYKELMAONQPKLSLPIRLLEPKDDAGLPPKATRGCRH 180
Db 121 ienakodlqlkvlvnsistqtehsykelmagngpklsiprllpekdaglppektrgrlh 180
XX
181 NCFDYSRCPLTSGFPVYVYSDQFVGSYLDPLVYKQAFQATARANYVTENADACLYVI 240
Db 181 ncfdysrcpltsqfpvyvysdqfvgssyldplvykqafqatvanyvtenaadacylvv 240
XX
241 LVGEMOEPVVLRAPELEKOLYSLPHWRTDGHNVYINLSKSPONTLLYVNSTGRAVAA 300
Db 241 lvgeompvvlrapelkolslphwrtldghnvynlslskspontllyvnstgravaa 300
XX
301 STEYVYVVRGFDLVVSPVHAMSEPNFMEIPPOVPYKRYLFTFOGEKTESLRSSLOEA 360
Db 301 steyvyvvrgefvlvsvplhamsepnfmeipovpykrylftfogekteslrssloea 360
XX
361 RSFEEME-----GDPP 372
Db 361 grgeeeqeklgvwrgrpp 357
XX
RESULT 7
AA84675
ID AAY84675 standard; Protein; 262 AA.

XX
XX AAY84675;
AC
XX 25-JUL-2000 (first entry)
DE
XX Amino acid sequence of a protein.
DE
XX Tumour necrosis factor receptor-associated Factor; TRAF;
KW TRF-protein-interacting hereditary multiple exostos protein; TREX;
KW signal modulator; tumour necrosis factor receptor;
KW CD40 mediated signal transduction; TRAF protein; cancer;
KW hereditary multiple exostos; autoimmune disease.
XX
XX Unidentified.
OS
XX WO200018959-A1.
XX
XX PD 06-APR-2000.
XX
XX 17-SEP-1999; 99WO-US21654.
PF
XX 17-SEP-1998; 98US-0156191.
XX
XX (UYCO) UNIV COLUMBIA NEW YORK.
XX
XX Sato T;
XX
XX WPI: 2000-293180/25.
XX
XX New nucleic acid encoding Tumour necrosis factor Receptor-Associated
PT factor (TRAF) protein-interacting hereditary multiple exostos (TREX)
PT protein, useful in the diagnosing cancer -
PS
XX Disclosure; Page 150; 161pp; English.
XX
XX The specification describes human and murine tumour necrosis factor
CC receptor-associated factor (TRAF) protein-interacting hereditary multiple
CC exostos (TREX) protein. TREX is a signal modulator which bridges
CC between the tumour necrosis factor (TNF) receptor and CD40 mediated
CC signal transduction. Anti-TREX antibodies are useful for treating an
CC abnormality in a patient by inhibiting binding of a TRAF protein and
CC a TRAF protein (that is, TRAF 2, TRAF 3 or TRAF 5). The abnormality
CC is cancer, a hereditary multiple exostos or an autoimmune disease. The
CC cancer is colon cancer, gastric cancer, human head and neck squamous
CC cell carcinoma, prostate carcinoma, breast cancer, thyroid cancer,
CC oesophageal cancer, lung cancer, colorectal cancer, ovarian cancer,
CC papillary bladder cancer, osteosarcoma, chondrosarcoma, liposarcoma,
CC giant cell tumour, Ewing sarcoma, or other malignant tumours. The
CC present sequence appears in the specification.
XX
XX Sequence 262 AA;
SQ
Query Match 29.5%; Score 1436; DB 21; Length 262;
Best Local Similarity 100.0%; Pred. No. 5.4e-128;
Matches 262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
658 VPPEOFTVWVLTREBERVNLNSLEPLNGLPYLKVVVWVMSPLPSBDLIMPDIYPIW 717
Db 1 vpregtfvwvltrebervnlnsleplnglpylkvvvwmvmsplpsbdlmpdiypiw 717
XX
718 VFTKXSLNRRFLPWNIEIETALISIDDAHLHDEIMFGFRVWRARPRIVFPGRYHA 777
Db 718 vftkxslnrrflpwnieietalisliddahlhdeimfgfrvwrarprivfpgryha 777
XX
61 vrtkxslmrlfipwneietealisliddahlhdeimfgfrvwrarprivfpgryha 120
Db 61 vrtkxslmrlfipwneietealisliddahlhdeimfgfrvwrarprivfpgryha 120
XX
778 WDIHQSWLYNSVNSVSCSLSVLTGAAPFHKYYAYLYSYVWPAQIRMDVDEYINCEDIA 837
Db 778 wdihqswlynsvnsvscslsvltgaapfhkyyaylysyvwpaqirmdvdeyincedia 837
XX
838 FLVSHITRRKPRIVVTRWMTFRGCGPQALSHDSHHERHCKLNFVVKYGVPLIXYTF 897
Db 838 flvshitrkprlvvtrwmtfrgcgpqalshdshherhcklnfvvkgyvplixytf 897
XX
181 flvshitrkprlvvtrwmtfrgcpqalshdshherhcklnfvvkgyvpliytf 240

QY 898 RVDSVLEFKTRLPDHTKCEKFI 919
DB 241 RVDGVLFKTRLPDHTKCEKFI 262

RESULT 8
ID AAY28883 standard; Protein: 718 AA.
AC AAY28883;
DT 01-FEB-2000 (first entry)
DE Human Ext-2 protein.
KW Human Ext-2 protein; Ttv protein; hedgehog interacting protein;
KW protein core; proteoglycan; hedgehog-mediated signal transduction;
KW hedgehog specific GAG chain; Ext dependant pathway; drug discovery assay;
KW growth factor; apoptosis; cell proliferation; cell differentiation; cancer.

OS Homo sapiens.
PN MO9950385-A2.
PD 07-OCT-1999.
PF 30-MAR-1999; 99WC-US06892.
PR 30-MAR-1998; 98US-0079928.
PA (HARD) HARVARD COLLEGE.
PI Bellaiche Y, The SI, Perlimon N;
DR WPI: 1999-610841/52.
DR N-PSDB; AA208224.
PT Identifying inhibitors of signal transduction of extracellular
PT proteins, used for developing agents for treating, e.g. proliferative
PT disorders or cancers or for cartilage -
PS Claim 9; Page 103-105; 14pp; English.
XX The present sequence is human Ext-2 protein. It is 25% identical to Ttv
XX protein. Exts are hedgehog interacting proteins which bind to hedgehog
XX proteins with high affinity to regulate diffusion. They are a family of
XX glycosyltransferases synthesizing GAG chains attached to the protein core
XX of proteoglycans. Hedgehog-mediated signal transduction can be inhibited
XX by inhibiting the synthesis of hedgehog specific GAG chains by Ext
XX dependant pathways. Drug discovery assays identifies such agents that
XX modify the biological activity of growth factors by inhibiting Ext
XX activity to prevent hedgehog-dependent proliferation of cells. This is
XX useful in treating disorders related to aberrant apoptosis, cell
XX proliferation or differentiation, cancer etc.
XX Sequence 718 AA;

Query Match 20.5%; Score 1000; DB 20; Length 718;
Best Local Similarity 32.5%; Pred. No. 1,1e-85;
Matches 260; Conservative 155; Mismatches 258; Indels 156; Gaps 27;

QY 140 HSKYELMAQNPRLSTL---PIRLPEKDDGLP-PRKATGCRLLHNCFDYSRCPITSGF-194
DB 48 hstlesndvkrstldvprvrlp-----adspipergdlscrmhtctfdyrc-----gfn 99
QY 195 -----PVYVYDSQGFV---FGSYLDPLVKQAFQATARA---NVYVTENADIACLYVILVGE 244
DB 100 pknltklylalkkyvdvdfgvsvantlsreyneilmalsdsdytdidnactlfvpsidv 159
QY 245 MQEYVVRPALELEKQLSLRHWRTDGHNIYIN-LSKRSOTQNLTVNSVIGRAMVQSTF 303

DB 160 lngn-trlketagamaqjstrw-drglnhllfmlpggppdyntaldvprdrallagggf 217
QY 304 YTVQYRPGFDLVVSPLVHAMSEPNFMEIPQVAVKRRYLTTFQGEKIESIRSSLOBARSP 363
DB 218 stwttyrgydvsl-pyysplsaee--vdlpekypgprqyfl-----lsqyqlhpey 265
QY 364 EEEEMGDPPADYDRIATLKAVODSKLDQVLEFCKQNPKPSLPTMALGGERE--DR 421
DB 266 redle-----alqvkhgeavlvldkctnlsegvlavwr-krchkhqvfdy 308
QY 422 LELLKSTFALITTPDPRIVISSGCATRLFEALENGAVPVVLEGOVQLPYOMLQNEA 461
DB 309 pylvgealtfcvllr--garl---ggav-lsdvlgagcvpvladsyilpfsevlwkra 361
QY 482 ALVVPKPRVVEVHFLRLSLSDSLAMRQGRFLMETYFSTADSIFFTVLAMIIRTRIQP 541
DB 362 svvvpkekmqdyvllqslprqleemqrgarvfweyfskatalatqlndrtlpy 421
QY 542 AADIREBAAEIRHSGKAAGTDPNMADGDLQGPVETEPYASPRYLNFLLVYTDY 601
DB 422 aalsyee-----wmdpav----- 435
QY 602 RSMNCAGPFRHLFPHTPDVLPSKFLGSGTGFRRIGSGAGSGCKEFOALGQVPRE 661
DB 436 -kwgsvsnpifl-----plppqs-----q 454
QY 662 QFTVVMILTYEREVEVLNMSLERLNGLPLNKKVVMVWNSP-KLPSDDLMPDIGVIMVVR 720
DB 455 gftaivltgydvslfvlvleevkypslskllvwmnqknpedalswplrvlplkvrt 514
QY 721 EKNSLNRRFLPMNEITEALISTDD-AHLRHDEIMGFVWRBARDRIVGPRGYHAWD 779
DB 515 aenklsnrffpydeleavlaiddlmltsdelqfygvwvrefpdrlyvgyrhlw 574
QY 780 IPHQSWLYNSNVCSELNVTGAAFFKXYAYLYSVYMPQAIQDMQDEVINCQDIAMNPL 839
DB 575 hemakwkyesewlnevsmyltgaafykhkyfnylytkmpgdlkhwdaImnncedlamnfl 634
QY 840 VSHTRKPRPIKVTSMRTFRQPCP--QALSHDSHFHERKICINFVKVYVPLLYTQF 897
DB 635 vanvtgkavlkvprkfkfpectaldglsldqtmwvresclnkfasvfgtmplkvwh 694
QY 898 RVDSVLEFKTRLPDHTKCF 916
DB 695 radpvlykddfp-eklsf 712

RESULT 9
ID AAB58946 standard; Protein: 663 AA.
AC AAB58946;
DT 27-MAR-2001 (first entry)
DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 654.
XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
XX neoplastic; neuroprotective; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; vulnerary; anticonvulsant;
XX antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
XX Addison's disease; allergy; autoimmune haemolytic anaemia;
XX autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
XX multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
XX cardiovascular disorder; wound healing; neurological disease.

OS Homo sapiens.
PN MO200055173-A1.
PD 21-SEP-2000.

PF 08-MAR-2000; 2000MO-US05881.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI: 2000-611515/58.
 N-PSDB; AAF21849.

PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -

Claim 11; Page 1098-1100; 1299p; English.

CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antitumor; vulnertary; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and agonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.

XX Sequence 663 AA;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 20.5%; Score 998; DB 21; Length 663;

Best Local Similarity 32.3%; Pred. No. 1,4e-85;

Matches 256; Conservative 127; Mismatches 252; Indels 158; Gaps 26;

QY 143 KELMAQNPRLSLPT-RLPEKDDAGLPPEKATRCRLHNCEDYSCRPRTSGF-----P 195

DB 4 reahitdvprvrlpadspilpergdl-----scrmhccfdvrc-----gfnprkxik 50

QY 196 VYVYDSQDFV-FGSYLDPLVKAQFQATARA---NVVTENADACLVILVGMQEPV 250

DB 51 vylalalkkyddfgvsvntistreyneellmaidsdylddlnacilfvsidvlnq-t 109

QY 251 LRPAELERQDLSLRHMTDGHNVIIIN-LSRKSPTOMLVNSTRAMVAOSFTYVQYR 309

DB 110 lrlketagamaqslrw-drgtnhllfmlp9gppdytaldvprdralll9ggfstvtyr 168

QY 310 PGFDLVSPVLMHSEPNFPIPOVVKRKYLFTEGKEIESIRSLQEARSEEMEG 369

DB 169 qgydvsl-pvysplsaee-vdlpekp9prqyfl-----lssvgjlhpeyrele- 215

QY 370 DPPADYDRIIATLKAQVDSKLDQVLEFCKNQPKPSLPTENALCGERE--DRLELKL 427

DB 216 -----alqvkhesvnlvldkcthlseglsvlr-krckkhvfdypqvlqe 259

QY 428 STEALITTPDPRVLISSGCAFLFEALLEVAVPVVLGEQVQLPYODMLQWNEAALVVPK 487

DB 260 atfcavlr-garl-----ggav-lsdvlgagcvpvvialdsyilpfisevldktrasyvpe 312

QY 488 PRTEVHFLIRSLSDSLAMRQGRFLVETYSADSTFNVLAMIRTRIQIPAAPIRE 547

DB 313 ekmsdvyslqslpqrqiemqrgarfwfweayfgskatalatclqjindrllyaaalsys 372

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

QY 548 EAAELPRHSGAAGTDPNMDNGDLGCVETEPPVYASPRYLRFNFTLTVDYFSNMCA 607
 DB 373 -----wddpav-----kwgsv 385
 QY 608 PCFHLFPHTPPDPVLPSPAKFLSGTGPRPFGGAGSGKRFQALGNNVRFQFTYVM 667
 DB 386 snplfl-----pllpqs-----qgftav 405
 QY 668 LTYREEVLMNSLERLNGLPYLNKVVVWNSP-KLPSEDLMPDQIGVIMVRFKNSLN 726
 DB 406 ltydvaelfrvitevskvpslklvwmnqknkppedsmpkrlrvklvrtaelnls 465
 QY 727 NRFLPMNLTEPAILSTDD-AHLKNDLIMGFRVWRARDRYGFPGRTYAMDIPHOSW 785
 DB 466 nrfpydeleaevaladdlmltsdelqfyewrefpdlvgyprlhlwdehmkw 525
 QY 786 LYSNYSCELSVLTGAFFPKRYAYLYLSYMPQAIRMVDEYINCEDIANNFLVSHTR 845
 DB 526 kysewelnvsmvltgaaftyhkylnlylytkmpgdikmwdaahmncediamflvanvtg 585
 QY 846 KPPIKVTSRWTFRCPCGP--QALSHDSHFHBRKCNIFEYKVGYPMLTYTQFRVSVL 903
 DB 586 kavlkvrprkkfkpectaigslldqthmversecinkfsavtvgtmpkvehndrpyl 645
 QY 904 FKTRLPMDXTGCF 916
 DB 646 ykddfp-eklxf 657

RESULT 10

ID AAW44851 standard; peptide; 728 AA.

XX AAW44851;

AC 20-AUG-1998 (first entry)

DE Hereditary multiple exostose; associated EXT2 gene isoform.

XX Hereditary multiple exostose; EXT2; chondrosarcoma; human; isoform;

XX treatment.

KW Homo sapiens.

XX EP837127-A2.

XX 22-APR-1998.

PD 26-AUG-1997; 97EP-0306503.

XX 21-OCT-1996; 96CN-0121928.

PR (UYHU-) UNIV HUMAN MEDICAL.

XX Deng HX, Fan CH, Xia J;

XX WPI: 1998-219110/20.

DR N-PSDB; AAV19375.

XX Cloned human EXT2 gene - associated with hereditary multiple

PT exostoses or chondrosarcoma

XX Claim 11; Pages 28-31; 31pp; English.

PS

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

This is the amino acid sequence of an isoform of the EXT2 gene associated with hereditary multiple exostoses and chondrosarcoma. The encoding polynucleotide is an isoform of the EXT2 gene described in Nature Genet., 14, 25, 1996. The polynucleotide can be used in the detection and treatment of EXT2-related diseases, and to identify compounds which activate or inhibit receptors for the encoded polypeptide. The polypeptide can be recombinantly produced by transforming or transfecting a cell with a vector containing the encoding nucleic acid.

Sequence 728 AA:
Query Match 19.3%: Score 939; DB 19; Length 728;
Best Local Similarity 31.6%: Pred. No. 7, 1e-80;
Matches 257; Conservative 126; Mismatches 256; Indels 174; Gaps 30;
140 HSKYKMAQNOPKLSL--PIRLPEKDAQLP-PPKATRCGLNCFDYSRCLTSGF-194
48 hslslnndwvkrslrdvrvrllp-----adslpergdlsrclhctcfdyrc-----gfn 99
195 -----PYVVDSDQFV--FGSYLDPLVKQAFQATARA---NVYTEMADIACTLVILGE 244
100 pknkikyiyialkkyvdffgvsanlsreynellmasdsdytddlntracflvpsldv 159
245 MOEPVVRPALEKOLYSLPMRTDGNHVIIN-LSRKSDONLLYVNSGRANVAOSTF 303
160 lqgn-clrlketagmqslrw-drgcnhllfmlp99ppdyntaldvprdrallag9gf 217
304 YTVQYRPGFDLVSPLVHANSEPNFMEIPQVPVKRYLFTFOGEKTESLSSLOEARSF 363
218 stwcyrgydvsl-pvysplsaee---vdlpekgyprgyfl-----lssgyqlpey 265
364 EEEMEGPPADYDRITATLKAQVDSKLDQYLVFTCKNQPKPSLPTMALCGERE--DR 421
266 redle-----alqvkhgsevlvldkctnlsegylsvr-krchkhqfydy 308
422 LELIKLSTFALIIRPGDPRLYISSGCATRLFEALEVGVAVPVLGEQVQLPYQDMLQWNEA 481
309 pqvlgaeftcfcvll--garl-----ggav-lsdvlgagcvprvldsyllpsevlwkra 361
482 ALVVPKPRVTEVHFLRLSLSDSLAMRRO-----GRFLMETVFSNADSIPTNL---- 531
362 svvvpkeemsdvylqlsqipqrgleemqrglfmeprvrrerw-----saanhqmalwpre 417
532 ----AMRTKQIQAARIPREAAAEIPIHRSKAGSTDPNADNDLDELPEVTEPPASP 587
418 qwdsqllndrtlypyaalsyee-----wmdpprav-- 445
588 RYLNENFTLTVIDFYSRNSNCAAGPRLHPHPPDVLPESEAKFLSGTGFRITGGAGSG 647
446 -----kqgsvanpfl-----pllppgs----- 463
648 KEFOALGAGNVRQGFVVMVLTREBEVYLNLSERLNGLPYLNKVVVVVWNSP-KLPSEDL 706
464 -----gftaivlygrveslffrvllevskvpslslkllvwmnqknknpeds 510
707 LMPDIGPIMVVRTEKNSLNRLFLPMNEIETFEALSTDDO-AHLRHDEIMGCFVWRRAR 765
511 lwpkltvplkvtrlaenklnrffrydeleaeavlaiddlmltsdelqfgyewreifr 570
766 DRIKPGFGRYHAMDIPHQSMLYNSNSCELSMVLTGAAFFHKYAYLYSYVMPOAIRDMV 825
571 drlvygprdlrlhwhemkkyesewcnevsmyltagaefykfyfnylytylmpgdikwv 630
826 DEYINCEDIAMNLTVSHITRRPKIKVTSRMTFRCPGCP--QALSHDSHFHRKHCINEF 883
631 dahmncedlamnflvanvgtkavlkvprkfkfkspecaidglsldegimvserseclnkf 690
884 KVUYGYMPLLYTQFRVDSVLEFKTRLPHDKTKCF 916
691 asvfgtmplkvhehradpvllykddfr-eklskf 722

XX Human Ext-1 protein; Ttv protein; hedgehog interacting protein;
KM tumour suppressor; multiple exostoses; diffusion regulation; GAG chain;
KM hedgehog protein; glycosyltransferase; protein core; proteoglycan;
KM hedgehog-mediated signal transduction; hedgehog specific GAG chain;
KM Ext dependent pathway; drug discovery assay; growth factor; apoptosis;
KM cell proliferation; cell differentiation; cancer.
XX Homo sapiens.
OS WO950385-A2.
PN 07-OCT-1999.
PD 30-MAR-1999; 99MO-US06892.
PF 30-MAR-1998; 98US-0079928.
PR (HARD) HARVARD COLLEGE.
PA Bellalche Y, The SI, Perrimon N;
PI WPI: 1999-610841/52.
DR N-PSDB: AA208223.
DR Identifying inhibitors of signal transduction of extracellular
PT proteins, used for developing agents for treating, e.g. proliferative
PT disorders or cancers or for cartilage -
PS Claim 9; Fig 3; 141pp; English.
XX The present sequence is human Ext-1 protein which is the putative tumour
CC suppressor for multiple exostoses. It is 56% identical to Ttv protein.
CC Extis are hedgehog interacting proteins which bind to hedgehog proteins
CC with high affinity to regulate diffusion. They are a family of
CC glycosyltransferases synthesising GAG chains attached to the protein core
CC of proteoglycans. Hedgehog-mediated signal transduction can be inhibited
CC by inhibiting the synthesis of hedgehog specific GAG chains by Ext
CC dependant pathways. Drug discovery assays identifies such agents that
CC modify the biological activity of growth factors by inhibiting Ext
CC activity to prevent hedgehog-dependent proliferation of cells. This is
CC useful in treating disorders related to aberrant apoptosis, cell
CC proliferation or differentiation, cancer etc.
XX
XX Sequence 746 AA:
Query Match 17.9%: Score 871; DB 20; Length 746;
Best Local Similarity 28.7%: Pred. No. 2, 3e-73;
Matches 236; Conservative 143; Mismatches 236; Indels 206; Gaps 33;
151 PKLSLPIR-LRP-----EKDAGLP-PPKATRG-----CRHNCFDYSRCLTSGF 194
56 prfepiriprtpwqqlenedsvshlsprkrdansslykqkcmesocfdtlck-kngf 114
195 PVVYV--DSQFVFGSYLDPLVKQAFQATARANYTEMADIACLYILVGEQEPVVL 252
115 kvuyypqkqgeklaeasy-----gnllaalegrsfytedpqaclfvslatldird-qls 167
253 PA-----ELEKOLYSLPMRTDGNHVIINL---SKSDONLLYVNSGRANVAOSTFTTV 306
168 pgyvhnlnrskvqslhw-nggrnllflnlysglvpdyvedvgfdl--gqamlaekasste 224
307 QYRPGFDLVSPV-----HAMSEPNFMEIPQVPVKRYLFTFOGEKTI-----ESLRS 356
225 nfrpofdvst-plfkskdhprtggergflkf-ntlprlkrkynlvfkqkylglsgdtrna 282
357 LQEARSEFEEMEGDPADYDRITATLK-----AVQDSKLDQYLVFTCKNQPKPSLPT 410
283 lyvhngedv-----lltckhgwkwqkhdksrdrndtey----- 319
411 EMALGGERDELLEIKLSTFALIIRPGDPRLYISSGCATRLFEALEVGVAVPVLGEQVQL 470

Db 320 -----ekydyemlnhnaiclv-prgrrrl-----gsfrfleaigaacypvmlnsngwel 366
 QY 471 PYODMLQNMNEALVYPRKPYTEVHFLLRSLSDDSLAMRGQFLMETPSTADSTFENVY 530
 Db 367 pfsvevnmqaaavgderlilqipsticrshqkllalrqtblweayfssvekylvlt 426
 QY 531 LAMITRTIOIPAAPIREBAALPIHRSRGAAGTDPNMDNGDLDPVETEPPEYAPRYL 590
 Db 427 leiigdrf-----fkhis----- 439
 QY 591 RNFITLVDPFRSMNCAPCPHILPPTPPDYLPEAKFLG-----SGTGRPTIGGAG 644
 Db 440 rnsli-----wnkhpq--glf-----vlpqysylgdfpyyanlglkp----- 476
 QY 645 GSGKEFOALGQNVREQTVMVLTFREREVLMSNLERLNGLPYLKVVVWVNSPK-LPS 703
 Db 477 --pskftavhavtp-----lvsqspvklilvaaksgycaqlilvncdplpa 525
 QY 704 EDLMPDIGVPIVAVVTEKNSLNRFLEPWEIETEALISIDDAHLRHDEIMGFRVWRE 763
 Db 526 kh-rwpatavpvyvlegesakvmsrflpydnalidaavlsidedtvlstevdfatwqs 584
 QY 764 AADRIVGPRGRYHNDIPHQSWLYNSNSCELSWVLGGAFFHKYAYLYVMPQAIRD 823
 Db 585 fperivgyparshvndakewgtskwtndysmvltgaaalyhkyhylyshylypaslkn 644
 QY 824 MVDEXINCEDIAMNPLVSHITRKPRIKYT-----SRMTFRCPGCPQALS 867
 Db 645 mvdcqlancedlmlnflvaavtklppikvltqkkykqyketmngqtsraswa----- 693
 QY 868 HDSHFHERHKCINFEVYVGYMPLLYTQFRVDSVLFKTRL 908
 Db 694 -dpdhfaqrsgcmntfawsfgympilhsgmldpvlfxdkqv 733

RESULT 12
 ID AAY84680 standard; Protein: 259 AA.
 XX
 AC AAY84680;
 XX
 DT 25-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of a protein.
 XX
 KW Tumour necrosis factor receptor-associated Factor; TRAF.
 KW TRF-protein-interacting hereditary multiple extoses protein; TREX;
 KW signal modulator; tumour necrosis factor receptor;
 KW CD40 mediated signal transduction; TRAF protein; cancer;
 KW hereditary multiple extosis; autoimmune disease.
 XX
 UN Unidentified.
 XX
 PN WO200018959-A1.
 XX
 PD 06-APR-2000.
 XX
 PF 17-SEP-1999; 99WO-US21654.
 XX
 PR 17-SEP-1998; 98US-0156191.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Sato T;
 XX
 DR WPI; 2000-293180/25.
 XX
 PT New nucleic acid encoding Tumour necrosis factor Receptor-Associated
 PT Factor (TRAF) protein-interacting hereditary multiple extoses (TREX)
 PT protein, useful in the diagnosing cancer -
 XX
 PS Disclosure; Page 155; 161pp; English.
 XX

CC The specification describes human and murine tumour necrosis factor
 CC receptor-associated factor (TRAF) protein-interacting hereditary multiple
 CC extoses (TREX) protein. TREX is a signal modulator which bridges
 CC between the tumour necrosis factor (TNF) receptor and CD40 mediated
 CC signal transduction. Anti-TREX antibodies are useful for treating an
 CC abnormality in a patient by inhibiting binding of a TREX protein and
 CC a TRAF protein (that is, TRAF 2, TRAF 3 or TRAF 5). The abnormality
 CC is cancer, a hereditary multiple extosis or an autoimmune disease. The
 CC cancer is colon cancer, gastric cancer, human head and neck squamous
 CC cell carcinoma, prostate carcinoma, breast cancer, thyroid cancer,
 CC oesophageal cancer, lung cancer, colorectal cancer, ovarian cancer,
 CC papillary bladder cancer, osteosarcoma, chondrosarcoma, liposarcoma,
 CC giant cell tumour, Ewing sarcoma, or other malignant tumours. The
 CC present sequence appears in the specification.
 CC
 SQ Sequence 259 AA;
 Query Match 17.4%; Score 848.5; DB 21; Length 259;
 Best Local Similarity 58.1%; Pred. No. 5.4e-72;
 Matches 151; Conservative 44; Mismatches 62; Indels 3; Gaps 2;
 QY 660 REQFTVMVLTFREREVLMSNLERLNGLPYLKVVVWVNSPKLPSSDLMPIGVPIVAVR 719
 Db 3 reqftvvltyerdaavlgaletlqglpynkllvwmvnrddpd-swpslhpvefir 61
 QY 720 TEKNSLNRFLEPWEIETEALISIDDAHLRHDEIMGFRVWREARDRIVGPRGRYHAMD 779
 Db 62 vaenlnrfypwdrleaeavlsiddidlmqgelilafvrerndrtivgprahha-- 119
 QY 780 IPHQSWLYNSNSCELSWVLGGAFFHKYAYLYVMPQAIRDMVDEXINCEDIAMNPL 839
 Db 120 rygdsfmlynshtcqmlllgaaltiknyltafyempaelrehvnsikncediamnyl 179
 QY 840 VSHITRKPRIKYTSRMTFRCPGCPQALSHDSHFHERHKCINFEVYVGYMPLLYTQFRV 899
 Db 180 vshltrkppiktsrwtlkcptcteslykgthfekrncmrlfcklygnplktsqtra 239
 QY 900 DSVLEFKTRLPHDKTKCFKEI 919
 Db 240 dsllfktrlpqnhqkcfkyv 259

RESULT 13
 ID ABB61749
 XX
 AC ABB61749;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide seq ID NO 12039.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 XX
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 N-PSDB; ABL05852.

THIS PAGE BLANK (USPTO)